THE AMERICAN
MATHEMATICAL MONTHLY

Volume 95, Number 1 January 1988

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Dept. of Mathematics, Stanford University, Stanford, CA 94305 (address effective until May, 1988). Shorter articles about the teaching of mathematics should be sent to STAN WAGON and JOAN HUTCHINSON, Department of Mathematics, Smith College, Northampton, MA 01063. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to ANITA E. SOLOW, Dept. of Mathematics, Grinnell College, Grinnell, IA 50112; in geometry or topology to DENNIS DETURCK, Dept. of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to RICHARD LIBERA, Mathematical Sciences Department, University of Delaware, Newark, DE 19717. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Proposed problems (three copies) and solutions (two copies), both elementary and advanced to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, University of Calgary, Alberta, Canada T2N 1N4.

Reprint Permission: A. B. WILLCOX, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: MS. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N W , Washington, DC 20036.

Back Issues: P. and H BLISS Co., Middletown, CT 06457.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W , Washington, DC 20036 and Montpelier, VT

The annual subscription price of $26 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $29 Student, unemployed and emeritus members receive a 50% discount, new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $90 per year

Second class postage paid at Washington, DC, and additional mailing offices

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Assocation of America, 1529 Eighteenth Street, N W , Washington, DC 20077-9564

PRINTED IN THE UNITED STATES OF AMERICA

Award for Distinguished Service to Professor Murray Klamkin

GERALD L. ALEXANDERSON

If you think of problems and of problem competitions, one person is sure to come quickly to mind: Murray Klamkin. In collections of problems, journal problem sections, or surveys of Putnam or Olympiad problems, Professor Klamkin's name is everywhere. For his contributions to the realm of problem solving, for his inspiring influence on young problem solvers, and for his many other contributions to mathematics, the Association presents this year's Award for Distinguished Service to Mathematics to Professor Murray S. Klamkin.

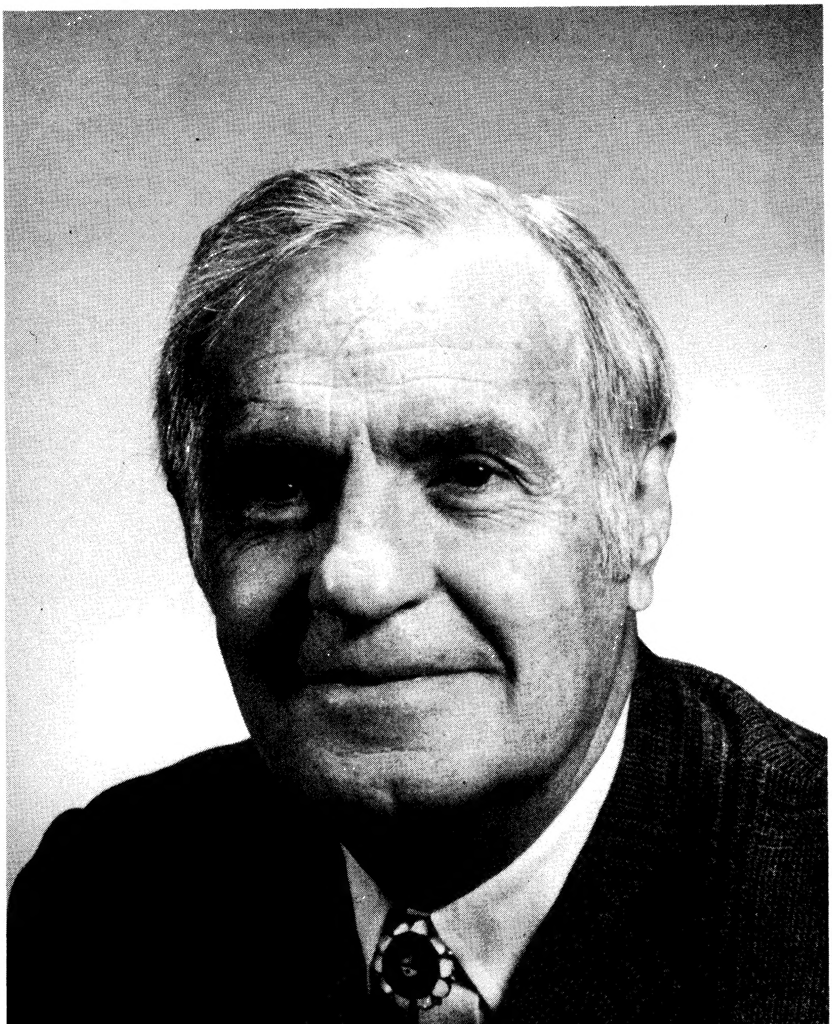
Professor Klamkin was born on March 5, 1921, in New York. He attended schools there, receiving a B. Ch.E. from the Cooper Union in 1942 and an M.Sc. in physics from the Polytechnic Institute of New York in 1947. Between 1948 and 1957 he moved from the rank of instructor to that of associate professor at the Polytechnic. From there he went on to a long and distinguished career in applied mathematics, first at AVCO Research and Advanced Development Division, then to a professorship at the State University of New York at Buffalo from 1962 to 1964. He then joined the Scientific Research Staff at the Ford Motor Company, 1965 to 1976. During that period he held a visiting professorship at the University of Minnesota and, after leaving Ford, he was appointed to a professorship in applied mathematics at the University of Waterloo. From 1976 until 1981 he was professor and chair of the Department of Mathematics at the University of Alberta, where he remains as professor of mathematics, now emeritus.

He has served as problem editor of many journals including *SIAM Review*, *The American Mathematical Monthly*, *Pi Mu Epsilon Journal*, *Mathematical Intelligencer*. He has edited the Olympiad Corner of *Crux Mathematicorum*, the Classroom Notes in Applied Mathematics of *SIAM Review*, and served as Associate Editor of *Computers and Mathematics with Applications* and of *Mathematical Modelling*.

Professor Klamkin has served the Association as a visiting lecturer between 1962 and 1970, as member of the Board of Governors from 1971 to 1974, and as member of numerous committees. In 1970 he became a member of the Putnam Prize Committee (the committee that each year makes up the problems for the William Lowell Putnam Mathematical Competition), a task for which he is ideally suited; he became chairman of that committee in 1972. The MAA presented him with a Certificate of Merit in 1978.

He has served other organizations as well including the American Association for the Advancement of Science of which he became a Fellow in 1982, the American Mathematical Society, and the Council of the Society for Industrial and Applied Mathematics. In 1987 the University of Waterloo conferred on him the degree, Doctor of Mathematics, *honoris causa*.

In recent years Professor Klamkin has been deeply involved in the Canadian, the U. S., and the International Olympiads. He has tirelessly coached the high school contestants, has contributed to the construction of competitions, and has written



Murray Klamkin

extensively on the Olympiads in *Crux Mathematicorum*. He served on the U. S. A. Mathematical Olympiad Committee as well as on its Canadian counterpart. He chaired the U. S. A. Committee between 1983 and 1985, coached the U. S. A. Team between 1975 and 1984, and coaches the mathematics team of university students on his own campus, simultaneously maintaining active involvement in mathematical competitions in the province of Alberta and in the city of Edmonton. He does all of this while carrying on extensive correspondence with a worldwide network of problem posers and problem solvers.

He has compiled and written solutions for the International Mathematical Olympiads 1978–1985 which appeared as volume 31 of the NML series in 1986, and he is performing the same task for the U. S. A. Mathematical Olympiads 1972–1986, also to appear soon in the NML.

Professor Klamkin has published over 150 papers in professional journals and has authored *1001 Problems in High School Mathematics*. Over the years he has given invited addresses around the world: in Sydney, Haifa, Suva, and Halifax, to name a few. He has almost certainly contributed problems to every mathematical journal with a problem section published in the English language, and to some that are not: *Elemente der Mathematik* and *Archief voor Wiskunde* among them.

Professor Klamkin's colleague, Andy Liu, has written: "I have greatly benefitted from my association with Murray Klamkin. He has generously shared with me many worthwhile and exciting experiences. What I value most are our common interest and close collaboration in problem solving. We both prefer working in the wee hours, with our telephones buzzing when some progress is made. His mathematical insight is uncanny."

And Professor Klamkin's longtime friend and collaborator, D. J. Newman has written: "Some of us remember when Mathematics was full of marvels and mystery. It was an adventure and a constant challenge. So it was with Murray Klamkin. A walk with Murray, something which I have often enjoyed, was always punctuated with a phrase such as, 'Yes, but is it true for tetrahedrons?'

"Whether or not we share his love for tetrahedra and such, very few of us could resist his excitement and love for the 'thing mathematical.'

"I do not want to give the impression that Murray Klamkin was just the mathematical 'enfant-terrible,' however. He is quite the scholar, in his own way, and can usually be counted on for the incisive reference, and the complete history behind a vast number of results."

"Of course his teaching, and especially coaching abilities are legend. I remember how delighted I always was when I heard his lectures or just his casual explanations."

We look forward to being challenged by Murray's ingenious problems for many years to come.

Isonemal Fabrics*

BRANKO GRÜNBAUM and G. C. SHEPHARD

BRANKO GRÜNBAUM and G. C. SHEPHARD teach mathematics at the University of Washington in Seattle, and the University of East Anglia in Norwich, respectively. They have published extensively on various questions in geometry, combinatorics, and group theory. During the last ten years they have worked on a systematic development of the theory of tilings and related topics; many of the results of this collaboration appear in the recently published book *Tilings and Patterns* (reviewed in this MONTHLY).

1. Introduction. The beginnings of geometry were stimulated by practical problems of building, surveying, and decorating. Throughout history, internally motivated developments as much as the needs of other fields helped steer geometry into many new directions, and endow it with a surprising variety of flavors. Foundations, algebraic geometry, and differential geometry are among the major and most widely studied branches of geometry; each is now a vast body of knowledge and insights. But geometry comprises a large number of other fields—smaller but no less interesting. Many of these disciplines (such as convexity, theory of symmetry, combinatorial geometry, computational geometry) arose, like geometry itself, from simple problems or observations, which often are visually and intuitively easy to grasp. Elementary beginnings often lead to deeper results, to applications in mathematics and outside of it, and to additional attractive problems; all this showing the versatility of the geometric approach, and its continuing vitality. As an added point of interest we may note that in many cases, the ease with which a problem may be formulated and understood is in marked contrast to the difficulties encountered in attempting its solution.

To illustrate this point, we find it surprising and even humbling to realize that we have no idea what is the maximum number of (two-dimensional) faces possible in a

*Research supported by NSF Grants MCS77-01629-A01, MCS8001570 and MCS8301971, and by a Fellowship from the John Simon Guggenheim Memorial Foundation.

convex polyhedron of which congruent copies will tile three-dimensional space—nor do we even known if any such maximum exists! Examples of "tiling polyhedra" with 38 faces were found by Engel **[8]**; additional information and references on this topic can be found in Grünbaum and Shephard **[16]**.

Many other questions arise (through some abstraction) from everyday phenomena and lead to similarly baffling problems; often enough, they also lead to interesting mathematical developments. In a number of papers we have discussed several topics of this nature—the classification of plane patterns, symmetries in ornamental art, tilings of space, colorings of patterns—but the field is still wide open. We hope that increasing numbers of our colleagues will be attracted to study these elegant and challenging questions.

In this article we are concerned with combinatorial, geometric, and group-theoretic problems related to **woven fabrics**. Two aspects of this topic may be as unexpected to the reader as they were to us. The first surprise was the existence of many nontrivial mathematical questions that arise in this context. The second surprise was our finding that over the centuries practical weavers have produced many fabric designs which are mathematically interesting. In **[15]**, **[19]**, and **[20]** we have given accounts of some of the questions and of the basic results concerning woven fabrics of the usual kind, in which the strands lie in two perpendicular directions; bibliographic references were also included, hence we do not repeat them here except as needed. In the present article we extend the investigation and consider fabrics in which the strands can lie in more than two directions. Some "fabrics" of these kinds have excellent mechanical features and have been produced since ancient times—not by weavers but by basket makers; see Section 8 for more details and for references. Recently we learned that a company concerned with the manufacture of textile machinery has produced a loom that weaves fabrics with strands in three directions (or "triaxial fabrics," as they call them), see **[41]**. We have also heard that such fabrics are being used in certain applications where non-stretchability is essential (as for sails or parachutes).

The organization of the article is as follows. In Section 2 we review the necessary terminology. Our treatment is self-contained but is necessarily brief; the reader is urged to consult **[15]** and **[19]** for additional information and illustrative examples. At the end of the section, the main result is formulated; the following four sections are devoted to its proof. In Section 7 we discuss various generalizations and open problems. Additional information on the practical aspects of the more general kinds of fabrics is given in the last section, together with some historical remarks.

Some of the fabrics we shall describe here are quite complicated, and verbal descriptions or even diagrams often prove inadequate for a detailed understanding of their structure. In order to help readers visualize some of the three-dimensional structures, in a number of cases (see Section 5) we have found it convenient to draw two parts of a fabric separately; the reader should fold a copy of the page so that one part lies on top of the other. We hope that this will indicate how the two parts are to be "interlaced." But for a full understanding of the material presented it is advisable to construct "models" of the fabrics. This can be easily done using long

narrow strips of colored paper to represent the strands; it will be found convenient to hold the strips in position by temporarily pinning or taping them to a suitable board.

2. Terminology and the statement of the main result. A fabric, as we use the term, consists of two or more sets of "strands" (or "threads") which are "woven" together subject to certain conditions. Our first aim will be to reformulate this statement as a rigorous mathematical definition.

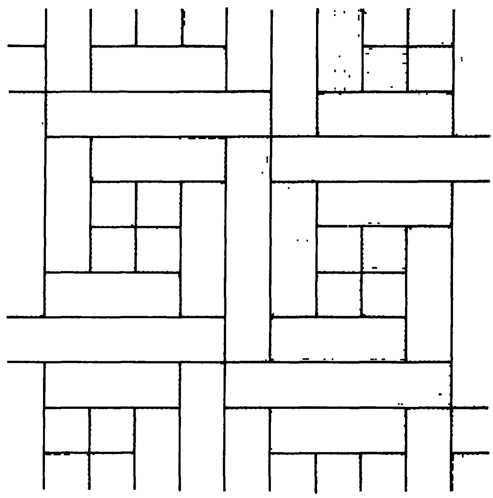
By a **strand** we mean a doubly infinite, straight, open strip of constant width, that is, the set of points of a plane that lie strictly between two parallel lines. A **layer** is a set of parallel, congruent and disjoint strands such that every point of the plane belongs to (the interior of) exactly one of the strands, or else lies on the boundaries of two adjacent strands (see [**15**, FIGURE 1]). Since all the strands in a layer are

B	*B*	*B*	*A*	*B*	*B*	*B*	*B*	*B*	*A*	*B*
A	*B*	*A*	*A*	*A*	*B*	*B*	*B*	*A*	*B*	*B*
B	*A*	*A*	*A*	*A*	*A*	*B*	*A*	*A*	*A*	*B*
B	*B*	*A*	*A*	*A*	*B*	*A*	*A*	*A*	*A*	*A*
B	*B*	*B*	*A*	*B*	*B*	*B*	*A*	*A*	*A*	*B*
B	*B*	*A*	*B*	*B*	*B*	*B*	*B*	*A*	*B*	*B*
B	*A*	*A*	*A*	*B*	*B*	*B*	*A*	*B*	*B*	*B*
A	*A*	*A*	*A*	*A*	*B*	*A*	*A*	*A*	*B*	*B*
B	*A*	*A*	*A*	*B*	*A*	*A*	*A*	*A*	*A*	*B*
B	*B*	*A*	*B*	*B*	*B*	*A*	*A*	*A*	*B*	*A*
B	*A*	*B*	*B*	*B*	*B*	*B*	*A*	*B*	*B*	*B*

2

1 A: (12) B: (21)

(a)



(b)

FIG. 1. A design (a) and a sketch (b) of a 2-way, 2-fold isonemal fabric; this is an example of a "sponge weave." Here and in the other illustrations we use A, B, etc. as abbreviations for the permutations which indicate the rankings of the strands.

parallel, there is a unique **direction** associated with the layer, and it is meaningful to speak of parallel or nonparallel layers. A **prefabric F** consists of two or more congruent layers of strands in the same plane E together with a preferential **ranking** or ordering of the layers at every point P of E that does not lie on the boundary of a strand. This ranking must satisfy the condition **R1** stated below, which formalizes the idea that the strands represent material objects which cannot penetrate each

other. The ranking is conveniently represented by a permutation of the numbers $1, 2, \ldots, f$, where f is the number of layers in the prefabric. Thus in a prefabric with three layers, the permutation (312) assigned to a point P means that at P we regard the strand belonging to layer 3 as taking a precedence over those belonging to layers 1 and 2, and the strand belonging to layer 1 as taking precedence over that belonging to layer 2. In accordance with the practical interpretation we say that, at P, the strand of layer 3 **passes over** the other two strands, or is the **uppermost** (or **top**) **strand**; the strand of layer 2 **passes under** the other two strands, or is the **lowermost** (or **bottom**) **strand**; and the strand of layer 1 **passes between** the other two strands, or is the **middle strand**. Obvious modifications of this terminology will be made when, for example, a prefabric contains more than three layers.

In order to come closer to the objects of practical use we shall in general restrict attention to **fabrics**, by which we mean prefabrics that satisfy the common-sense requirement **R2** expressing the idea that a fabric should "hang together."

The restrictions on the rankings are as follows.

R1. Suppose that the point P belongs to a strand T_i of layer i and also to a strand T_j of layer j (where $i \neq j$). Then if layer i is ranked before layer j at P, layer i must be ranked before layer j at *every* point of $T_i \cap T_j$.

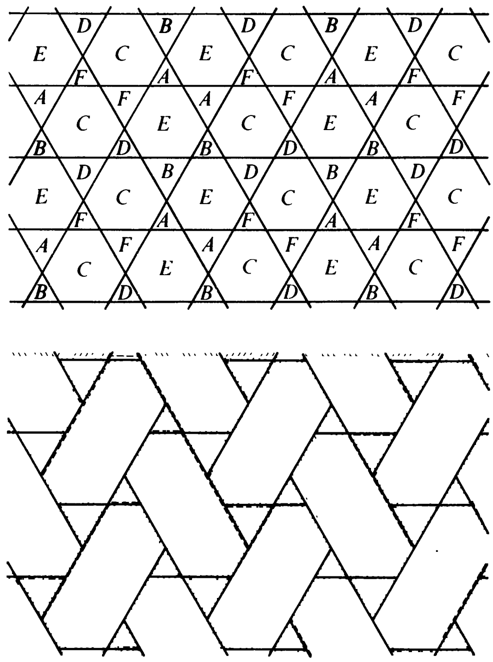
R2. It is impossible to partition the set of all strands, belonging to all the layers, into two nonempty subsets so that each strand in the first subset passes over (is ranked before, or takes precedence over) every strand in the second subset.

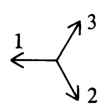
A prefabric that fails to satisfy **R2** is said to "fall apart." Examples of such prefabrics with two layers were given in [**15**, FIGURE 5] and [**19**, FIGURE 2]. An example of a prefabric with three layers will be given shortly (FIGURE 4). Criteria for testing whether a given prefabric satisfies condition **R2** were given by Clapham [**2**], [**3**] and [**4**]; see also Enns [**9**]. For a practical procedure see [**19**]; although the formulation there is for prefabrics with just two layers, it is of general applicability.

If the strands of a fabric **F** lie in w different directions, then **F** will be called a **w-way fabric**, and if **F** contains f layers then **F** will be called an **f-fold fabric**. (It is clear that these definitions make sense for prefabrics as well; if convenient, we shall use them in this more general sense.) If no further restrictions are imposed on the fabrics under consideration, it is easy to see that w-way, f-fold fabrics exist for all integer values of w and f satisfying $f \geq w \geq 2$. In [**15**], [**19**], and [**20**] we considered the "minimal" case $w = f = 2$; here we shall be mainly concerned with larger values of w and f.

In order to specify a w-way, f-fold fabric we shall employ a graphical representation that generalizes the traditional designs of black and white squares (introduced by weavers long ago, and used in most mathematical treatments as well). On the plane E of the fabric **F** we use sets of equidistant parallel lines to represent the edges (boundaries) of the strands; a w-way fabric leads to lines lying in w different directions. These lines divide E into a set of polygonal regions or **tiles**, each of which is labelled with a permutation indicating the ranking of the layers at one

(and, therefore, by **R1** at every) point of the region. The resulting labelled tiling is called a **design** for the fabric. Examples of designs are given in FIGURES 1, 2, 3 and subsequent diagrams. FIGURE 1 shows a 2-way, 2-fold fabric known traditionally as a **sponge weave**. FIGURES 2 and 3 show 3-way, 3-fold fabrics, and they illustrate the fact that if $w \geq 3$, the relative positions of the strands can differ in such a way as to lead to infinitely many different tilings of the plane.





A: (123) B: (231) C: (312)
D: (132) E: (213) F: (321)

FIG. 2. A design and a sketch of a 3-way, 3-fold fabric. This fabric is not isonemal. The positions of the layers can be moved relative to one another to form a continuous family of distinct fabrics.

Many of the illustrations show, besides the designs, also **sketches** of the fabrics. In these, shading is used to indicate for each tile the layer to which the uppermost strand belongs. Hence a sketch approximates the "appearance" of a model of the fabric when it is looked at from above. It should be noted that, unlike a design which specifies the structure of a fabric completely, a sketch gives much less information about the fabric when $f \geq 3$. This is illustrated by FIGURE 7; the fabrics whose designs are shown in FIGURES 7(a) and 7(c) are distinct yet their sketches are the same.

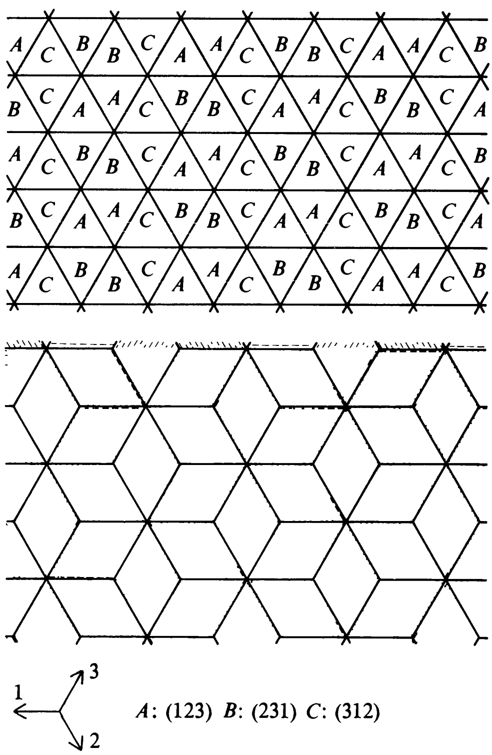


FIG. 3. A design and a sketch of a 3-way, 3-fold isonemal fabric known as the "mad weave."

Given any tiling defined by sets of parallel and equidistant straight lines, various restrictions on the allocation of permutations to the tiles (implied by conditions **R1** and **R2**) have to be satisfied if the resulting labelled tiling is to be the design of a fabric. In particular, as a consequence of condition **R1**, if the common edge of two adjacent tiles S_1 and S_2 is parallel to the strands of layer i, then the permutations allocated to S_1 and S_2 can differ only in the position of the integer i. We also note that this implies, in the case of w-way fabrics with $w \geq 3$, that the permutation allocated to any one tile is uniquely determined by those allocated to all its adjacent tiles. "Designs" which violate condition **R2** are more difficult to recognize on sight. An example is shown in FIGURES 4(a) and 4(b). Here the "design" does not correspond to a fabric since the strands labelled * in the diagram pass over all the remaining strands. In other words, a prefabric constructed according to this "design" would separate into the two parts shown in FIGURES 4(c) and 4(d). See Section 7 for further comments on the topic of such "weaves."

A **symmetry** of a fabric **F** is defined to be any isometry of the plane E containing **F** which maps each strand of **F** onto a strand of **F** and either preserves the rankings at each point of E, or else reverses all the rankings. A symmetry which reverses the rankings is said to **interchange the sides** of **F**, and one that leaves the rankings

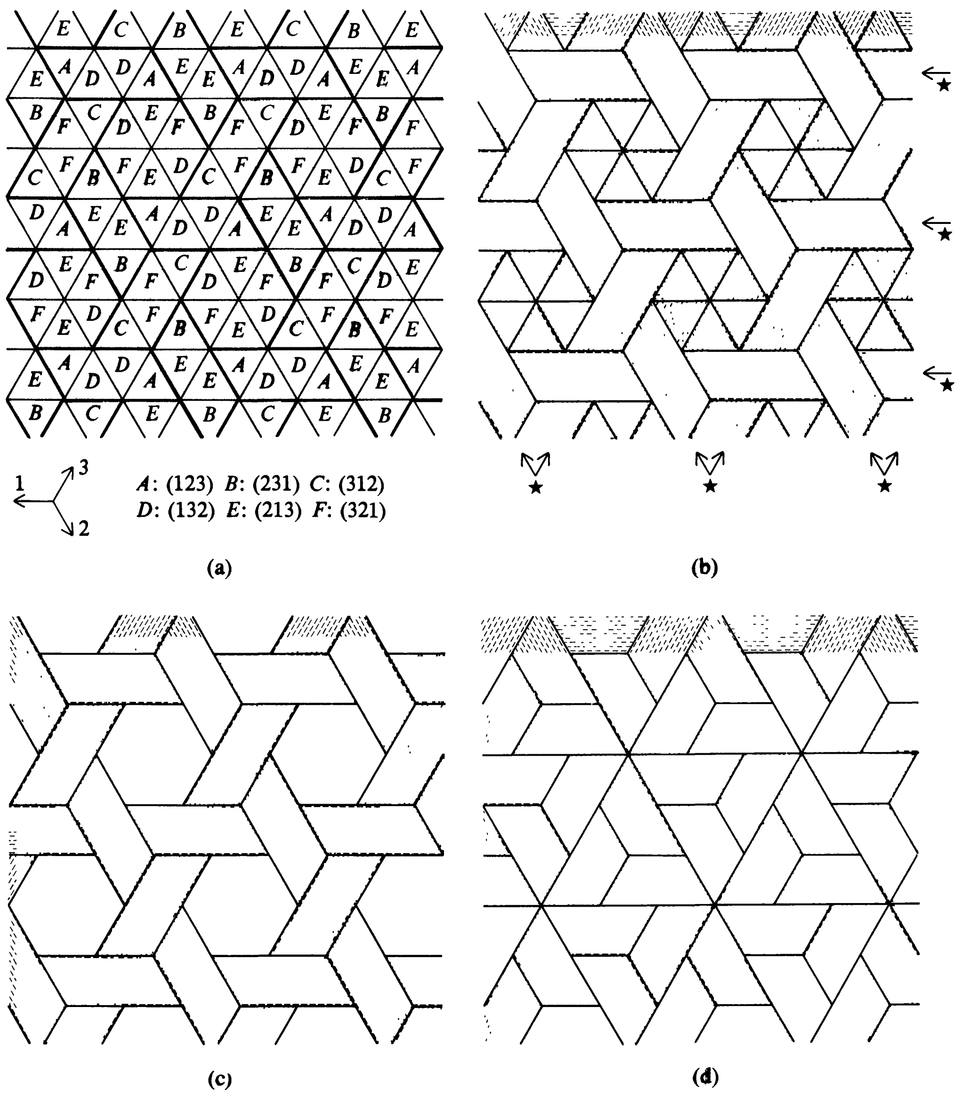


FIG. 4. The design in part (a) appears to be that of a fabric, but is not since it "falls apart," that is, violates condition **R2**. In the sketch (b) we have indicated by asterisks a subset of the strands that pass over all the remaining strands and so can be "lifted off." In (c) and (d) we show sketches of the resulting pseudofabrics.

unchanged is said to **preserve the sides** of **F**. The group of all symmetries of **F** (with the operation of composition) is denoted by **S(F)** and is called the **symmetry group** of **F**. We use $\mathbf{S_0(F)}$ to denote the normal subgroup of **S(F)** consisting of those symmetries that preserve the sides of **F**. There are two possibilities: either $\mathbf{S_0(F)} = \mathbf{S(F)}$, or $\mathbf{S_0(F)}$ is of index 2 in **S(F)**. In the latter case there must be some element σ in **S(F)** which interchanges the sides of the fabric, and then $\mathbf{S(F)} = \mathbf{S_0(F)} \cup \sigma\mathbf{S_0(F)}$.

A fabric is called **periodic** if $\mathbf{S}_0(\mathbf{F})$, and, therefore, $\mathbf{S}(\mathbf{F})$, contains translations in at least two nonparallel directions. The design of a periodic fabric can be obtained from a **fundamental block** (usually a parallelogram called a **period parallelogram**, but for groups with 3-fold rotations frequently chosen as a regular hexagon) by applying to it the translations of $\mathbf{S}(\mathbf{F})$.

A fabric $\mathbf{F}$ is called **isonemal** (pronounced i-so-nee′-mal, from the Greek word for thread) if $\mathbf{S}(\mathbf{F})$ is transitive on the strands of $\mathbf{F}$. Thus an isonemal fabric is one with the property that every strand is the image of one chosen strand under a symmetry of the fabric or, equivalently, the strands form one orbit under the group $\mathbf{S}(\mathbf{F})$. For example, the 3-way, 3-fold fabric of FIGURE 3 is isonemal—in this case even the proper subgroup $\mathbf{S}_0(\mathbf{F})$ of $\mathbf{S}(\mathbf{F})$ is transitive on the strands. On the other hand the 3-way, 3-fold fabric of FIGURE 2 is not isonemal.

Throughout the discussion, we consider two fabrics as identical if one can be mapped onto the other by a similarity transformation, possibly combined with an interchange of the two sides of the fabric (that is, a reversal of all the rankings in the design of the fabric). Clearly, fabrics with the general appearance of that in FIGURE 2 but corresponding to different ratios of size of the small triangles to the large ones, are different under this interpretation. In the case of isonemal fabrics, to say that two fabrics are identical can be shown to be equivalent to saying that they are **homeomeric**. A coarser classification, into **henomeric** types, is also possible and seems to be the appropriate one in some cases. (The meaning of these terms is, in essence, that two objects are homeomeric if there is a homeomorphism mapping one onto the other and compatible with all the symmetries involved; henomeric is defined similarly, but using a bijection between parts of the object instead of the homeomorphism. For detailed explanations see [**17**] or [**21**], where examples of applications of these concepts to many classes of geometric structures are given as well.)

Using the terminology introduced above we are now able to state the basic result:

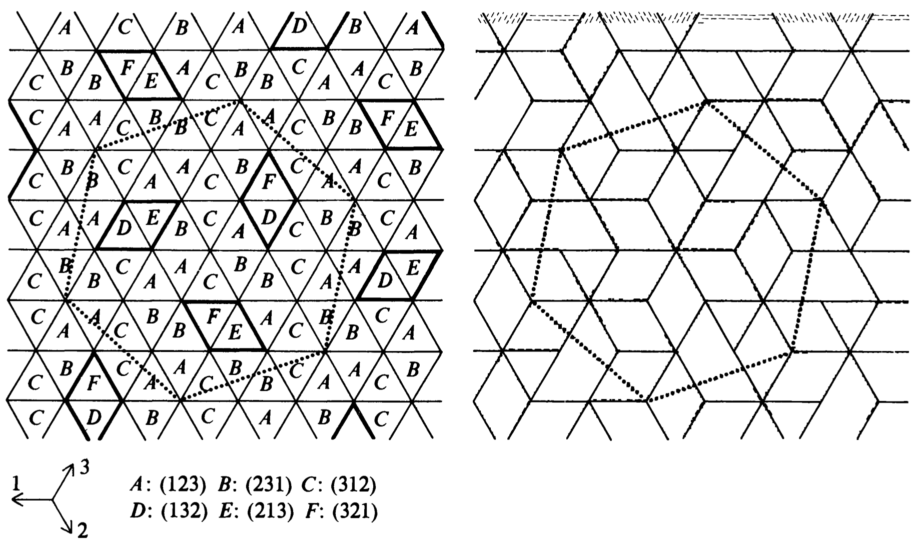
THEOREM. *If w and f are such that a w-way, f-fold periodic isonemal fabric exists, then the pair (w, f) is one of the following six*: $(2,2)$, $(2,4)$, $(4,4)$, $(3,3)$, $(3,6)$ *or* $(6,6)$. *Moreover, for each of these pairs there exist infinitely many different w-way, f-fold periodic isonemal fabrics.*

The proof of the theorem falls into two parts. First we shall prove the last (existential) assertion. We already know that infinitely many different 2-way, 2-fold isonemal fabrics exist (see [**15**]), and the next three sections will establish the corresponding facts for each of the other five pairs (w, f). In Section 6 we shall complete the proof by showing that no pairs (w, f) can occur other than the six listed in the statement of the theorem. In Sections 3 through 6 all the isonemal fabrics are periodic, and we shall not repeat this each time.

3. Construction of 3-way, 3-fold isonemal fabrics. It is clear that if a 3-way fabric is to be isonemal, the directions of the layers must be equally inclined (at 60° or 120° to each other). Hence the edges of the strands must be arranged as illustrated

in FIGURES 2 or 3. However, it is easy to verify that the situation in FIGURE 2, where the edges of the tiles do not pass completely across strands, is incompatible with isonemality. Thus only those fabrics can be isonemal in which some points (the centers of 3-fold rotational symmetry) are on the boundaries of strands of all three directions.

The fabric of FIGURE 3 is the "simplest" 3-way, 3-fold isonemal fabric, in the sense that it is periodic with a fundamental block of the smallest area. This fabric is known as the "mad weave," see Section 8. Infinitely many other isonemal fabrics can be derived from the "mad weave" by the procedure of transposing the rankings of pairs of strands in suitable parts of the fabric. If this is done in a sufficiently "orderly" way, isonemality will be preserved. An example is shown in FIGURE 5. In each of the rhombs outlined by a heavy line in the design, the rankings of two strands have been interchanged (compare FIGURE 3). It is clear that infinitely many fabrics of this kind can be constructed in a similar manner.



(a) (b)

FIG. 5. The construction of a 3-way, 3-fold isonemal fabric by transpositions. Starting from the "mad weave" (FIGURE 3), we transpose two of the strands in each of the rhombs indicated by thickened lines in the design (a). If such transpositions are carried out in a sufficiently orderly way, as done here, the resulting fabric is isonemal. A hexagonal fundamental block is indicated by dotted lines. A sketch of this fabric is shown in (b).

We shall now describe some further families of 3-way, 3-fold isonemal fabrics, in order to illustrate the wealth of possibilities. In FIGURE 6 we show two members of the first family, which we propose to call **hexagonal fabrics**. Let us denote by $\mathbf{F_h(p)}$

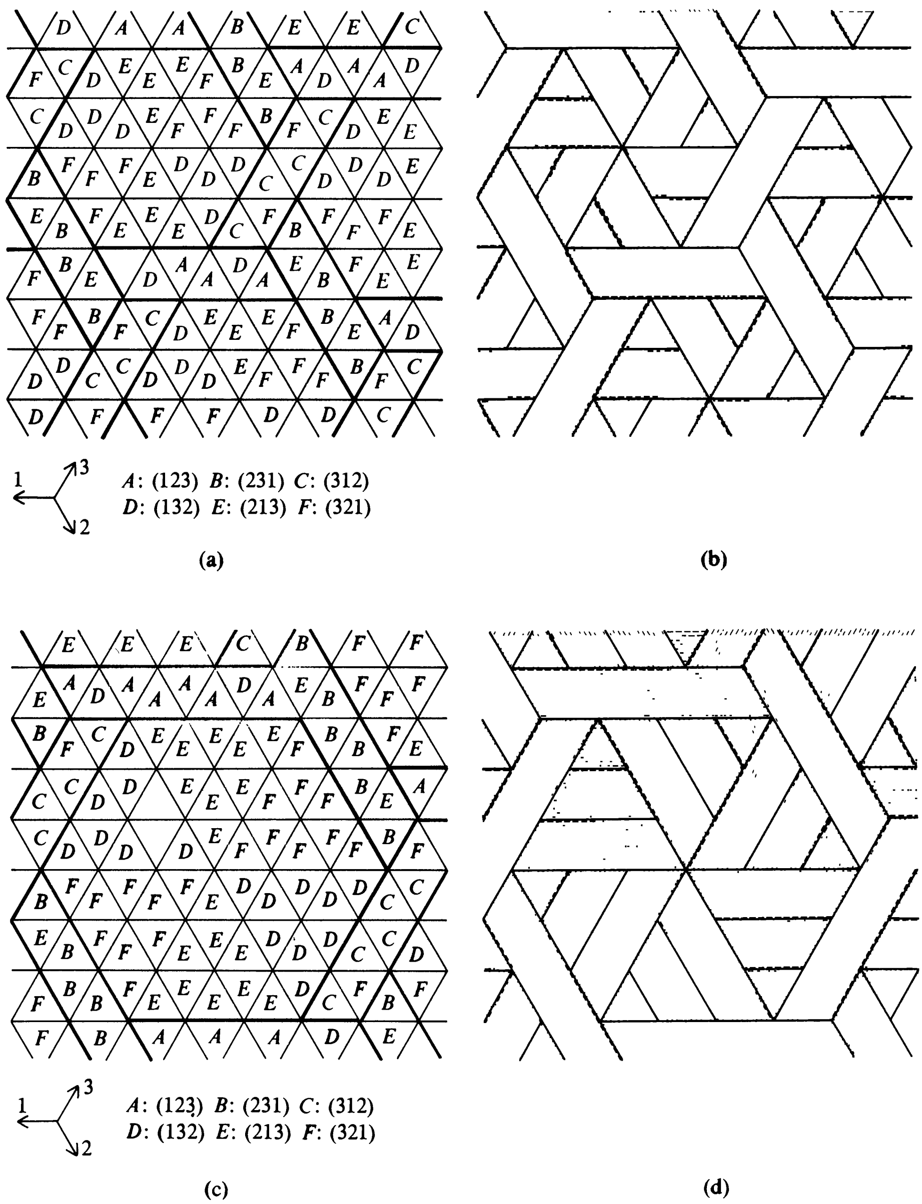
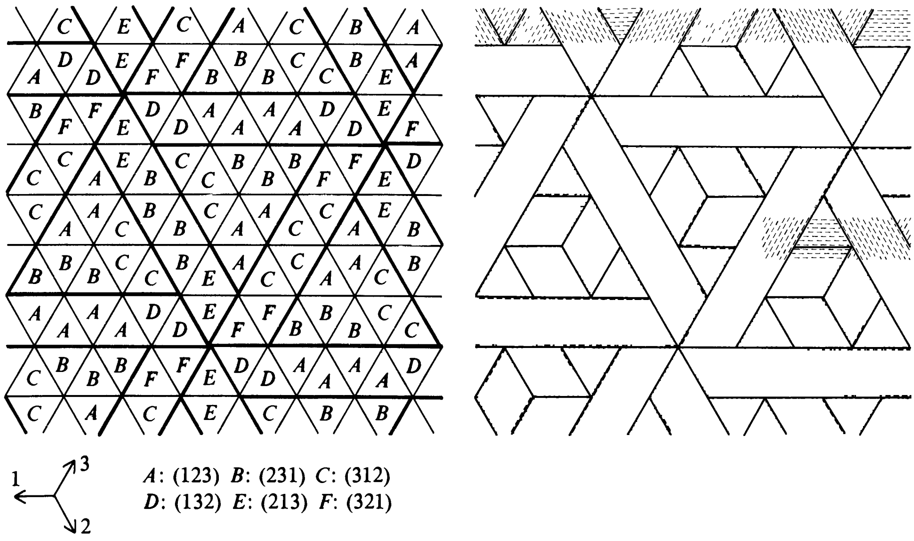
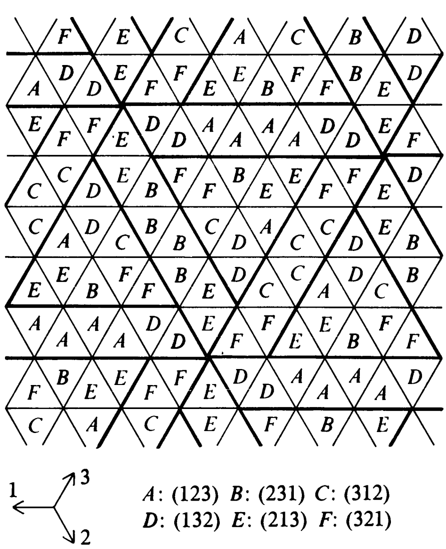


FIG. 6. Designs and sketches of two 3-way, 3-fold isonemal fabrics for which we propose the name "hexagonal fabrics." Parts (a) and (b) show $\mathbf{F_h(2)}$ and parts (c) and (d) show $\mathbf{F_h(3)}$. These are the first two members of an infinite family of isonemal fabrics denoted $\mathbf{F_h(p)}$ with $p \not\equiv 1 \pmod 3$. In the design (and in several of the following ones) some of the lines are thickened to facilitate comparison between the design and the corresponding sketch.

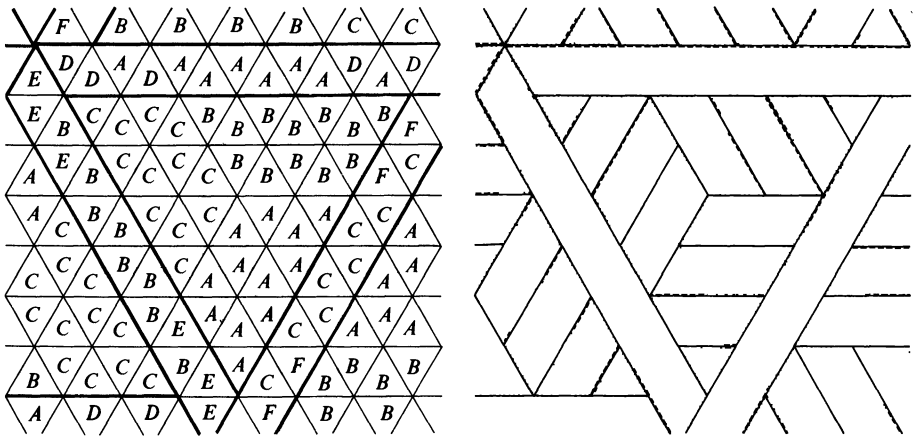


(a) (b)



(c)

FIG. 7. Designs and sketches of two 3-way, 3-fold isonemal fabrics for which we propose the name "triangular fabrics." Parts (a) and (b) show $\mathbf{F_t(1)}$ and parts (d) and (e) show $\mathbf{F_t(2)}$. The triangular fabrics $\mathbf{F_t(p)}$ (with $p \geqslant 1$) form an infinite family of isonemal fabrics. Part (c) shows the design of another 3-way, 3-fold isonemal fabric, which is obtained from $\mathbf{F_t(1)}$ by transposing certain pairs of strands. A sketch of this fabric is identical with that shown in (b), thus illustrating the inadequacy of sketches in specifying w-way fabrics when $w \geqslant 3$.



A: (123) B: (231) C: (312)
D: (132) E: (213) F: (321)

(d) (e)

FIG. 7.

the hexagonal fabric in which the hexagonal "patches" are each built up from six equilateral triangles which appear in the sketch as made up of p parallel strands. The hexagonal patches are separated by **floating strands**, that is, by strands which run for an appreciable distance along one side of the fabric without interweaving with strands from the other two layers. The prefabric of FIGURE 4 appears to belong to this family, but as already mentioned, it is not a fabric since it does not hang together. In general, it is not hard to see that for $p > 1$ the prefabrics $\mathbf{F_h(p)}$ are indeed fabrics, and that they are isonemal if and only if $p \not\equiv 1 \pmod 3$.

In FIGURES 7(a) and 7(c) we show designs of two members of a second family of 3-way, 3-fold fabrics. These we call **triangular fabrics**, and we denote $\mathbf{F_t(p)}$ the triangular fabric in which the floating strands separate triangular patches each of edge-length $3p$. It can be shown that $\mathbf{F_t(p)}$ is an isonemal fabric for each $p \geq 1$. The case $p = 0$, in which the triangles disappear completely, can be interpreted as meaning the "mad weave" of FIGURE 3. The "mad weave" may therefore be denoted by $\mathbf{F_t(0)}$; alternatively, with equal justification it may be denoted by $\mathbf{F_h(0)}$.

Hexagonal and triangular fabrics can be modified and combined in various ways. An example is shown in FIGURES 8(a) and 8(b). Here the hexagonal patches have edge-length 2, the triangular patches have edge-length 3, and they are separated by floating strands as before. More generally, it can be shown that if a fabric of this kind has hexagonal patches of edge-length p and triangular patches of edge-length q, then it is isonemal if and only if $p \not\equiv 1 \pmod 3$ and the integers $2p + 1$ and $2q + 1$ are relatively prime. Another example of a fabric which results from

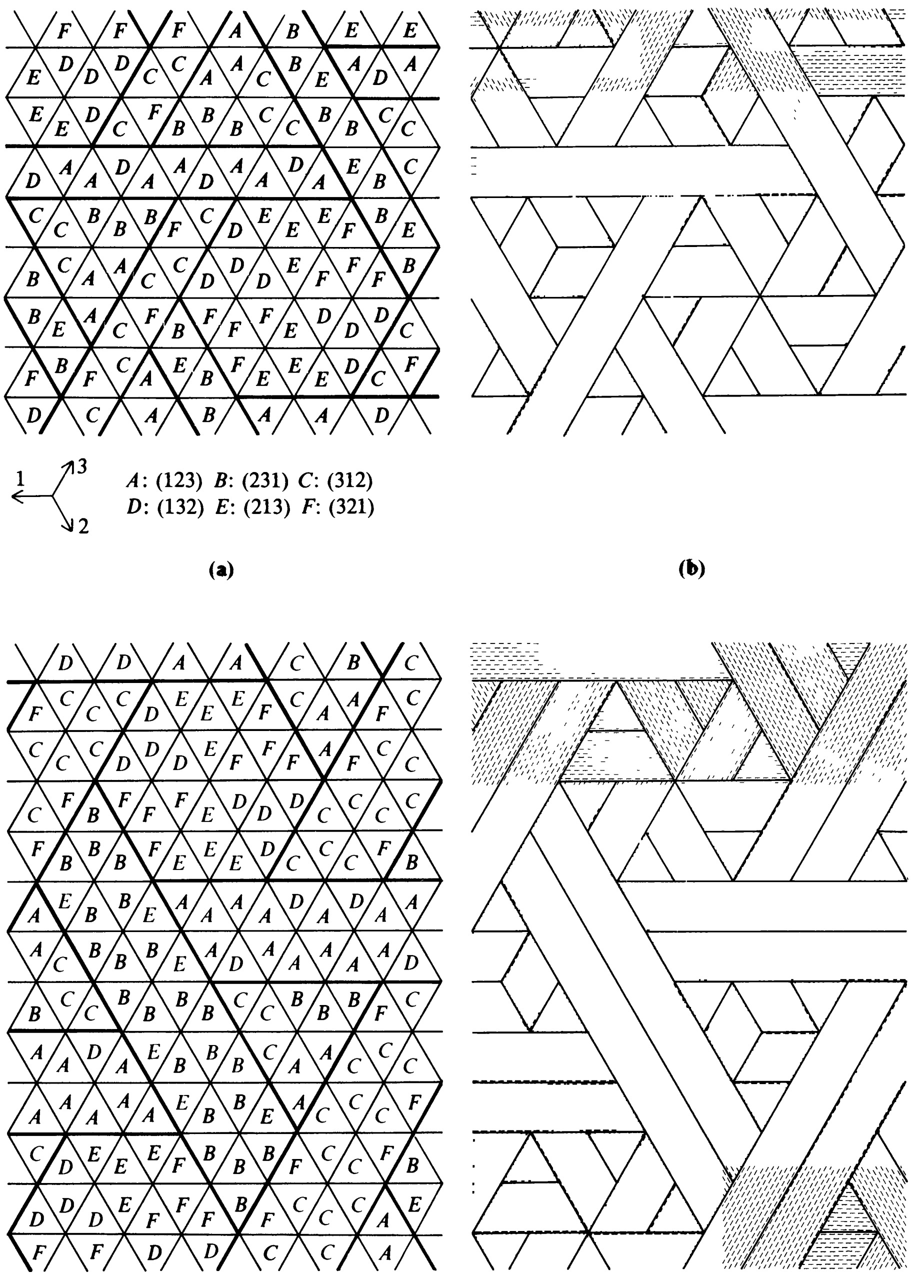


FIG. 8. Designs and sketches of two 3-way, 3-fold isonemal fabrics that result from combining the triangular and hexagonal fabrics shown in the previous two figures.

combining a hexagonal and a triangular fabric is shown in FIGURES 8(c) and 8(d). Here the patches are separated by pairs of floating strands. It is clear that there are many other possibilities.

4. Construction of 2-way, 4-fold and 3-way, 6-fold isonemal fabrics. The simplest way to construct a w-way, $2w$-fold fabric $\mathbf{F}^*$ ($w = 2$ or 3) is by taking any w-way, w-fold fabric $\mathbf{F}$ and replacing every strand by a pair of strands whose boundaries coincide (in other words, by two strands one of which lies on top of the other). If $\mathbf{F}$ is such that

(i) the group $\mathbf{S}_0(\mathbf{F})$ of symmetries that preserve the sides of the fabric is transitive on the strands, and

(ii) the group $\mathbf{S}(\mathbf{F})$ contains a symmetry which interchanges the sides of the fabric,

then the corresponding fabric $\mathbf{F}^*$ will be isonemal. It is clear how a design for $\mathbf{F}^*$ can be derived from that of $\mathbf{F}$, and the sketches of the two fabrics will be identical.

More interesting isonemal fabrics can be obtained by a procedure we call **interlacing**. In practical weaver's terminology this is known as "stitching" or "tying" the two fabrics together, see [**13**, Chapter 7]. It can be explained as follows. Let $\mathbf{F}$ be a w-way, w-fold fabric ($w = 2$ or 3) and lay one copy of $\mathbf{F}$, say $\mathbf{F}_1$, on another copy of $\mathbf{F}$, say $\mathbf{F}_2$, facing in the other direction and in such a way that the boundaries of the strands of $\mathbf{F}_1$ and of $\mathbf{F}_2$ coincide. Then clearly the prefabric $\mathbf{F}_1 \cup \mathbf{F}_2$ is not a fabric since it does not hang together, but under certain circumstances it may be transformed into a fabric (even an isonemal fabric) by reversing the ranking of the top strand of $\mathbf{F}_2$ and the bottom strand of $\mathbf{F}_1$ at various points of the plane. If $\mathbf{S}_0(\mathbf{F})$ is transitive on the strands, and the reversal of rankings or interlacing is carried out in some "orderly" way, then an isonemal fabric may be obtained. Examples starting from 2-way, 2-fold fabrics are shown in FIGURE 9. In FIGURE 9(a) we begin with two copies of the common "plain" or "tabby" weave (one with strands in layers 1 and 2, the other with strands in layers 3 and 4) and interlace the strands at the positions indicated by heavy lines in the design. In FIGURE 9(b) we show the result of applying a similar procedure to the fabric traditionally known as the "basket weave", and in FIGURE 9(c) to a $\frac{3}{3}$ twill. (For an explanation of these terms see [**15**].) All of these fabrics are isonemal.

Exactly similar considerations apply to 3-way, 3-fold fabrics. We give an example in FIGURE 10. Here two copies of the "mad weave" of FIGURE 3 are placed one on top of the other, and interlacing takes place at all points inside the hexagons indicated by heavy lines. The precise method of doing this is shown by the permutations in the design.

The construction of fabrics obtained by interlacing is not completely trivial, since it is often difficult to determine, from a given design, when interlacing is or is not possible and if it leads to an isonemal fabric. In all the examples given here the results have been carefully checked, theoretically and by using models as suggested in the introductory remarks. The "doubling up" of all strands, described at the start

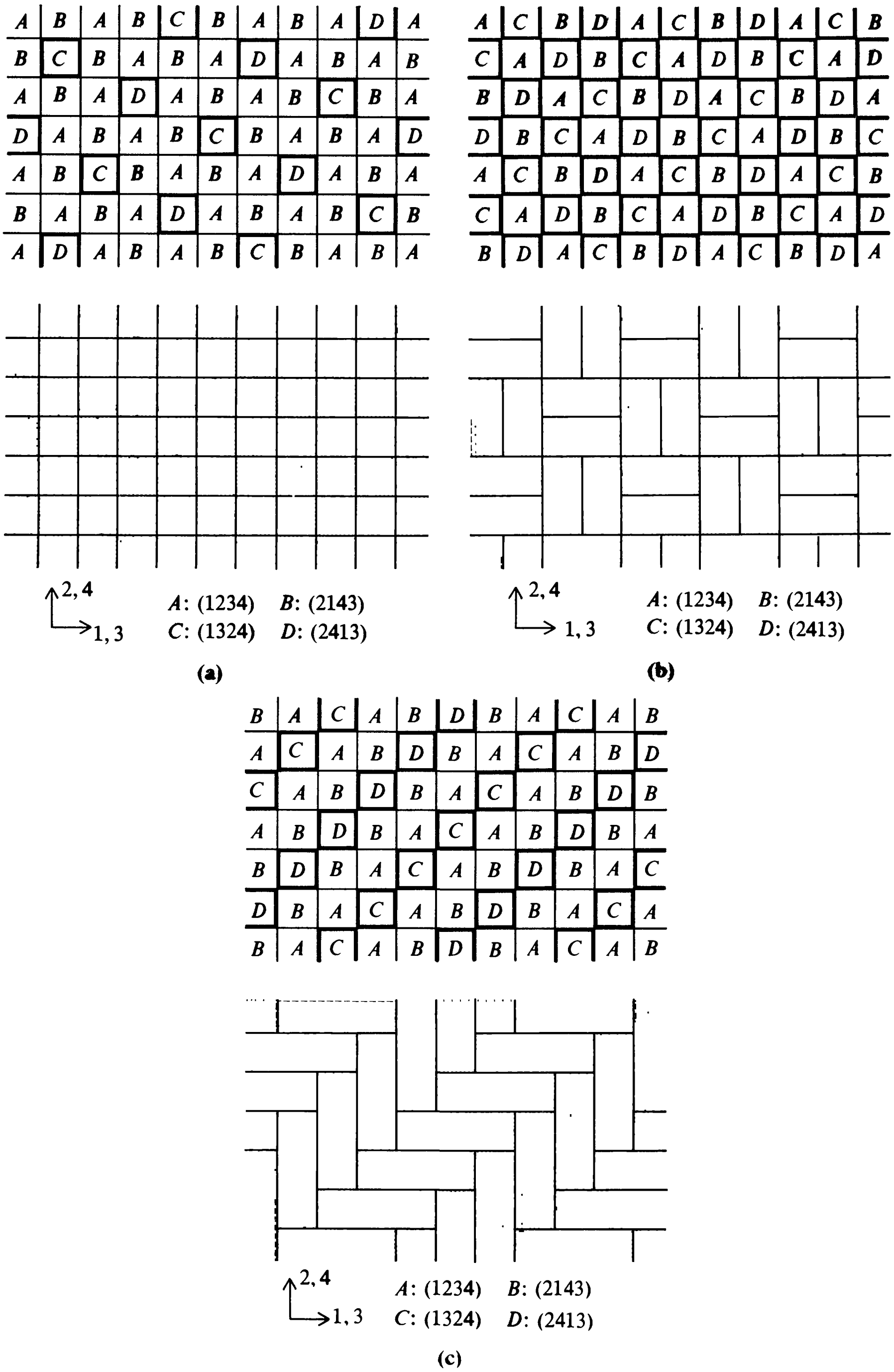
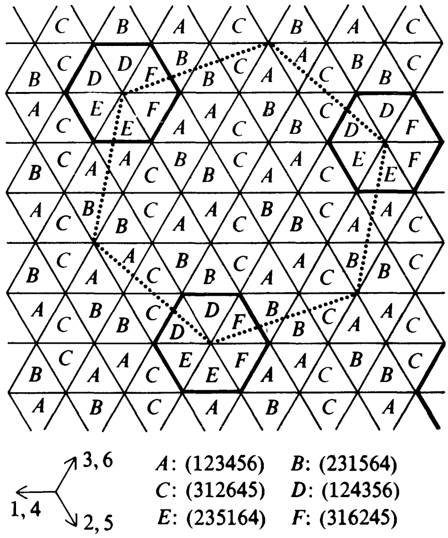


FIG. 9. Designs and sketches of three 2-way, 4-fold isonemal fabrics that are obtained from 2-way, 2-fold fabrics by the process of "interlacing." In part (a) we show how two copies of the "plain" or "tabby" weave can be laid one on top of the other so that the boundaries of the strands coincide, and then interlaced at the positions shown by the thickened squares in the design. In (b) we show a similar process applied to a "basket weave," and in (c) to a twill. It should be noticed that the sketch of each fabric coincides with that of each of its component 2-way, 2-fold parts.



A: (123456) B: (231564)
C: (312645) D: (124356)
E: (235164) F: (316245)

FIG. 10. Interlacing two copies of the "mad weave" (FIGURE 3) produces a 3-way, 6-fold isonemal fabric **F**. The sketch of this fabric coincides with that of the "mad weave," and the thickened hexagons indicate the places where the strands are interlaced. The dotted hexagon indicates a fundamental block of the fabric **F**.

of this section, can obviously be considered as an extreme special case of interlacing.

It should be noticed that the method of interlacing can also be used to construct different kinds of 2-way, 4-fold isonemal fabrics, for which the boundaries of the strands in the two copies of **F** are distinct (that is, parallel strands *partly* overlap). The practical uses of 2-way, 4-fold fabrics appear in most cases to have the strands of the two layers shifted a half-width with respect to each other; see, for example [**13**, Figure 7.11], for such an isonemal fabric obtained by interlacing two copies of a $\frac{2\ 1}{2\ 1}$ twill. (Due to differences in the methods of presentation used, the drawing of this fabric in [**13**] resembles a sketch of a 2-fold fabric.)

On the other hand, it can be shown that in the case of 3-way, 6-fold fabrics, the interlacing procedure can yield isonemal fabrics only if the boundaries of the two copies of **F** coincide.

5. Construction of 4-way, 4-fold and 6-way, 6-fold isonemal fabrics. The only method known to us of constructing these fabrics is by a variation of the interlacing procedure described in the previous section. For $w = 2$ or 3, we place two copies of a w-way, w-fold fabric **F** together as in Section 4, but in such a way that the corresponding layers of the two copies are *not parallel.* If we interlace the copies so that they hang together then we obtain a $2w$-way, $2w$-fold fabric, and it may be possible to arrange for this to be isonemal.

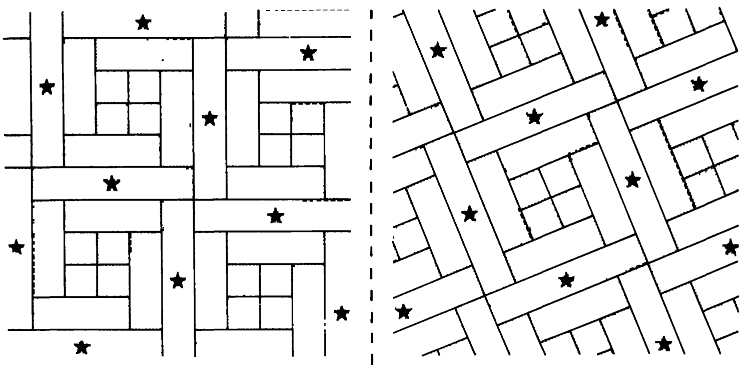


FIG. 11. A 4-way, 4-fold isonemal fabric. Two copies of a "sponge weave" (FIGURE 1) are laid one on top of the other in such a way that the strands indicated by asterisks can be interlaced. The reader is invited to fold the page along the dashed line, to see that then the asterisks in the two copies of the "sponge weave" coincide. This representation has been chosen since a design for the fabric would be so complicated as to be unintelligible.

Fabrics produced in this way are difficult to describe since their designs are usually so complicated as to be unintelligible, and sketches do not suffice. We, therefore, adopt the method indicated in the examples of FIGURES 11 and 12. In FIGURE 11 we show sketches of two copies of the "sponge weave" of FIGURE 1, and indicate how the page is to be folded so that the two sketches lie one on top of the other. If this is done carefully, it will be noted that the long floating strands marked with asterisks cross each other in such a way that they can be interlaced. If each such pair is interlaced then it is easily verified that the resulting 4-way, 4-fold fabric is isonemal. In FIGURE 12 we give a second example. Here a 6-way, 6-fold isonemal fabric is obtained in a similar manner by interlacing two copies of the 3-way, 3-fold hexagonal fabric $\mathbf{F}_h(2)$ of FIGURE 6.

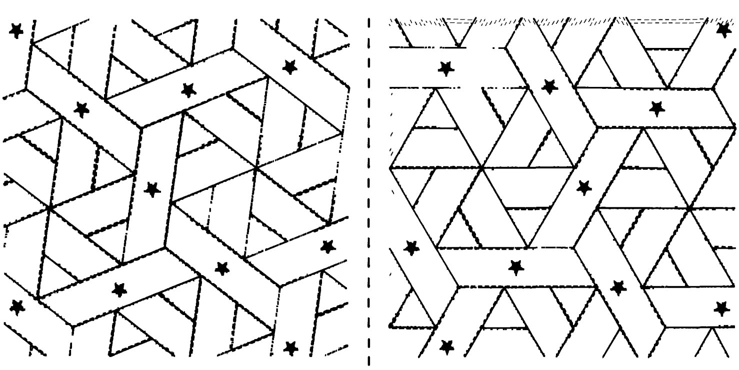


FIG. 12. A 6-way, 6-fold isonemal fabric obtained by interlacing two copies of the hexagonal fabric $\mathbf{F}_h(2)$ (see FIGURES 6(a) and (b)). The representation is the same as in the previous figure.

The same procedure may be applied to many 2-way, 2-fold fabrics with sufficiently long floating strands, such as other "sponge weaves" (see [**15**, Figure 6] for examples), and to many 3-way, 3-fold fabrics such as the hexagonal and triangular fabrics described in Section 3. Other fabrics of these kinds may be obtained by interlacing in FIGURES 11 and 12 only suitable proper subsets of the strands marked by asterisks. We deduce that infinitely many 4-way, 4-fold and 6-way, 6-fold isonemal fabrics exist, as claimed.

6. Characterization of the (w, f) pairs. From [**15**] we know that infinitely many 2-way, 2-fold isonemal fabrics exist, and in the previous three sections we have established the existence of infinitely many w-way, f-fold fabrics where (w, f) is one of the pairs $(2,4)$, $(4,4)$, $(3,3)$, $(3,6)$, or $(6,6)$. To complete the proof of the theorem it is, therefore, only necessary to show that no other pairs (w, f) can occur for isonemal fabrics.

We shall accomplish this by considering the properties of the group $\mathbf{S}_0(\mathbf{F})$; this group must be one of the 17 crystallographic plane groups (for a description of these groups see, for example, [**6**], [**10**], [**21**], [**30**], or [**38**]).

If $\mathbf{S}_0(\mathbf{F})$ contains a reflection in a line L (lying in the plane of the fabric) then every strand of $\mathbf{F}$ must be either parallel to L or else perpendicular to it. Hence, if $\mathbf{S}_0(\mathbf{F})$ contains lines of reflection lying in more than one direction, then these lines must be in two perpendicular directions. In any case, whether all the lines of reflection are parallel or not, all images of any given strand T of $\mathbf{F}$ under the group $\mathbf{S}_0(\mathbf{F})$ lie in either one or two directions.

If $\mathbf{S}_0(\mathbf{F})$ contains no reflections but contains glides (glide-reflections), then $\mathbf{S}_0(\mathbf{F})$ must be either the group pg (in which all the glides are along parallel axes) or the group pgg (in which the glides are along axes in two mutually perpendicular sets of lines parallel to each other). In either case all images of any given strand T of $\mathbf{F}$ under the group $\mathbf{S}_0(\mathbf{F})$ must lie in either one or two directions. It should be noted that in the latter case the two directions do not have to be perpendicular; see, for example [**15**, Figure 22].

If $\mathbf{S}_0(\mathbf{F})$ contains neither reflections nor glides then it must be one of the groups $p1$, $p2$, $p3$, $p4$, or $p6$. It follows from this that all images of a given strand T of $\mathbf{F}$ under the symmetries of $\mathbf{S}_0(\mathbf{F})$ must lie in 1, 1, 3, 2, or 3 directions, respectively.

We deduce from the above that in all cases the images of T under $\mathbf{S}_0(\mathbf{F})$ must lie in 1, 2, or 3 directions. We now ask in how many directions will the images of T lie when we apply the whole symmetry group $\mathbf{S}(\mathbf{F})$ instead of $\mathbf{S}_0(\mathbf{F})$? The question is trivial if $\mathbf{S}(\mathbf{F}) = \mathbf{S}_0(\mathbf{F})$; if the two groups are not equal, the fact that $\mathbf{S}_0(\mathbf{F})$ is a normal subgroup of index 2 in $\mathbf{S}(\mathbf{F})$ shows that the number of directions either remains the same or doubles. The latter occurs when the image of a strand T under *some* symmetry in $\mathbf{S}(\mathbf{F})$ is not parallel to the image of T under *any* symmetry in $\mathbf{S}_0(\mathbf{F})$. We deduce that every isonemal fabric is w-way for $w = 2$, 3, 4, or 6; the value $w = 1$ is clearly inadmissible.

We now consider the possible values of f for an isonemal fabric to be f-fold. If all the strands visible on the upper side of $\mathbf{F}$ lie in w directions (where $w = 2$ or 3)

then so must all the strands visible on the lower side of **F**. The latter may be in different directions from those on the upper side, or they may be in the same directions but distinct from them, or they even may be the same strands. These three possibilities lead (for $w = 2$) to 4-way, 4-fold; 2-way, 4-fold; or 2-way, 2-fold fabrics. Or (for $w = 3$), to 6-way, 6-fold; 3-way, 6-fold; or 3-way, 3-fold fabrics, respectively.

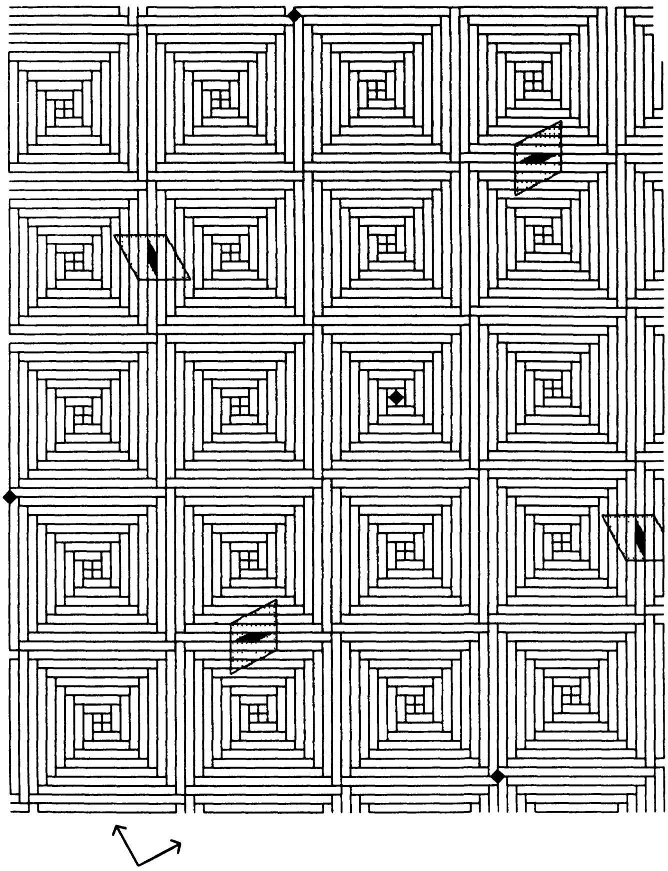


FIG. 13. A sketch of a 4-way, 4-fold isonemal fabric which illustrates the fact that it is possible for strands in all four directions to be visible on one side of the fabric. The construction is similar to that in FIGURE 11. Here two copies of a "sponge weave" with long "floating strands" are used; the upper copy is shown and the lower has strands in the directions indicated by the arrows under the diagram. The two copies are placed in such a way that sets of five adjacent parallel strands from the lower fabric can be interlaced with sets of five adjacent parallel strands from the upper fabric; this is not visible from the top, but is indicated by the shaded areas in the sketch. The middle strand of each of these interlaces can then be interlaced again—those from the bottom fabric with the topmost layer of the upper fabric as indicated by black rhombs. In the same way, strands from the upper fabric are interlaced with the lowermost layer of the bottom fabric. It is easily verified that the fabric so produced is a 4-way, 4-fold isonemal fabric with strands of all four directions visible on each side of the fabric. The small black squares mark centers of 4-fold rotational symmetries of the resulting fabric.

In the remaining case, the strands visible on the upper side of **F** lie in $2w$ directions (where $w = 2$ or 3). Now the strands visible on the lower side of **F** must be the same as those visible on the upper side. This follows since, by the earlier arguments, half the strands visible on the upper side are images of a strand T under symmetries of **S(F)** which do not belong to $\mathbf{S}_0(\mathbf{F})$, and which, therefore, interchange the sides of the fabric. Hence, we obtain, in this case, either a 4-way, 4-fold fabric (for $w = 2$), or a 6-way, 6-fold fabric (for $w = 3$).

This completes the characterization of the possible values of w and f given by the Theorem.

The possibility discussed at the end of this proof can actually occur. This is illustrated by the 4-way, 4-fold fabric sketched in FIGURE 13; analogous 6-way, 6-fold fabrics can also be constructed. It will be noted that the fundamental block of the fabric in FIGURE 13 is exceedingly large; it is not known whether such fabrics exist with reasonably small fundamental blocks.

7. Comments. (a) Clearly the symmetry group of an isonemal fabric must contain translations. If these translations lie in at least two nonparallel directions then we have called the fabric *periodic*, and our main theorem relates to such fabrics. However, there exist isonemal fabrics in which all the translations are parallel, and we shall refer to these as *nonperiodic*. A sketch of such a fabric appears in FIGURE 14. In fact, the theorem remains valid even for nonperiodic isonemal fabrics since it can be shown that every nonperiodic isonemal fabric must be either

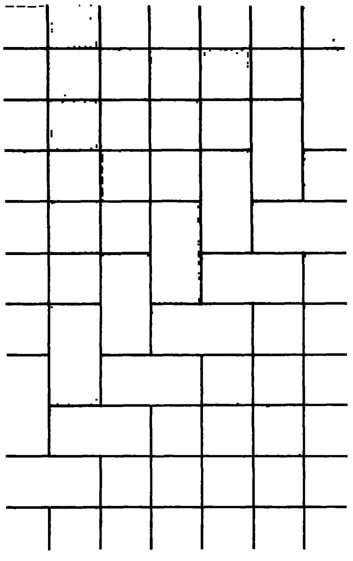


FIG. 14. A sketch of a nonperiodic 2-way, 2-fold isonemal fabric. The diagonal "stripe" continues in both directions, and the rest of the fabric is the same as in the "tabby weave." The only translational symmetries are parallel to the diagonal, hence the fabric is not periodic according to the definitions adopted. A 2-way, 4-fold isonemal fabric can be obtained, for example, by replacing each strand by two strands whose boundaries coincide.

2-way, 2-fold or 2-way, 4-fold. The sketch in FIGURE 14 may be regarded as representing a fabric of either kind—for the 4-fold fabric the strands in both directions should be doubled, as explained at the beginning of Section 4.

(b) A fabric **F** is said to be **strongly mononemal** if, in the design **D** of **F**, every strip that corresponds to a strand of **F** is an isometric image of every other strip, under an isometry that either preserves all the rankings or else reverses them all. (The isometry in question is not required to be a symmetry of the design **D**.) We conjecture that our Theorem remains valid for strongly mononemal fabrics.

A weaker restriction applies if the fabric **F** is **weakly mononemal**. For that, in the design **D** of **F**, every strip that corresponds to a strand T of **F** must again be an isometric image of the strip that corresponds to any other strand T^*. However, the isometry that accomplishes this is not required to preserve (or reverse) all the rankings, only the position of the layer to which the strands T and T^* belong. In other words, if T is on top in one tile, then T^* should be on top of the corresponding tile regardless of the rankings of the other strands in these tiles.

Clearly, for 2-way fabrics the concepts of weakly and strongly mononemal fabrics coincide. It is not known whether for 3-way fabrics either of the two mononemality properties implies isonemality.

It seems likely that the Theorem remains valid even for weakly mononemal fabrics.

(c) In Section 4 it was shown how 2-way, 4-fold and 3-way, 6-fold isonemal fabrics can be obtained by interlacing two 2-way, 2-fold or 3-way, 3-fold fabrics. Using the criteria for a fabric to hang together given in **[3]** or **[26]**, it is easy to show that, conversely, every 2-way, 4-fold isonemal fabric can be obtained from two 2-way, 2-fold fabrics by interlacing. It is not known whether an analogous converse is valid for 3-way, 6-fold; 4-way, 4-fold; and 6-way, 6-fold isonemal fabrics, or in the case that all the fabrics involved are assumed to be (strongly, or else weakly) mononemal.

(d) It seems likely that for *isonemal* 3-way prefabrics condition **R2** is automatically fulfilled. It even may be that the same is true for (strongly or weakly) mononemal prefabrics. If either of these possibilities is valid, then it would easily follow that all (isonemal, or mononemal) 3-way, 6-fold fabrics arise by interlacing.

(e) The concept of a fabric can be generalized in various ways. For example, we may modify the definition of a layer of strands by deleting the condition that every point of the plane belongs to the closure of a strand. In other words, we now allow the possibility that the (congruent and parallel) strands of a layer may be separated by gaps. A "fabric," defined exactly as in Section 2 but using these modified layers, will be called a **pseudofabric** or a **weave**. Examples of pseudofabrics appear in FIGURES 4(c), 4(d), 15, and 16. (The weave of FIGURE 4(c) is frequently used in basketry, especially in the manufacture of chair seats. It is one of the weaves that can be woven mechanically, see **[41]**.) It is clear that if **F** is a 2-way pseudofabric then it is possible to obtain a fabric $\mathbf{F}^*$ by widening the strands of **F**; moreover, if **F** is isonemal, it is possible to do this in such a way that $\mathbf{F}^*$ is isonemal as well. Such replacement of w-way pseudofabrics by fabrics is clearly not always possible if $w \geq 3$; see, for example, the weave in FIGURE 4(c).

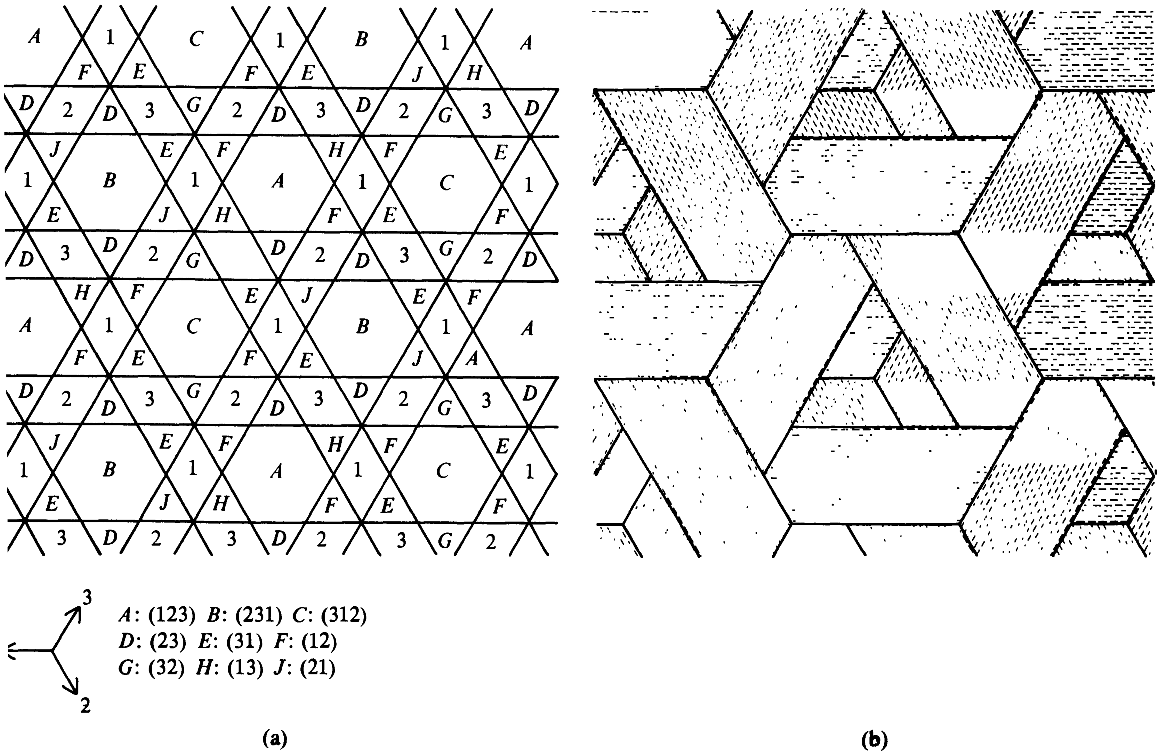


FIG. 15. A design (a) and sketch (b) of a 3-way isonemal pseudofabric. In each of the three directions the strands are separated by a gap equal in width to half that of the strand.

A simple but interesting 3-way, 3-fold pseudofabric is shown in FIGURE 16. We have not found it mentioned anywhere, although similar—but not isonemal—weaves are of common use in basketry.

We believe that mathematically meaningful results on pseudofabrics are possible, and present here some relevant ideas.

Starting from a given isonemal fabric, it is sometimes possible to produce a pseudofabric by omitting every alternate strand (or every $k \geq 1$ consecutive strands out of every $k + 1$ or $k + 2$) in every layer. Examples are easy to find, but we have been unable to formulate necessary and sufficient conditions for this procedure to produce an isonemal weave.

For any point P of the plane let $s(P)$ denote the number of strands of a weave **F** that contain P. Let $f = \max s(P)$, the maximum being over all points P of the plane which are not on the boundary of any strand. Then the weave **F** is said to be **f-fold**. Most of the other definitions and properties of fabrics carry over to pseudofabrics without any modifications. For example, an **isonemal pseudofabric** may be defined as one whose symmetry group is transitive on its strands. The

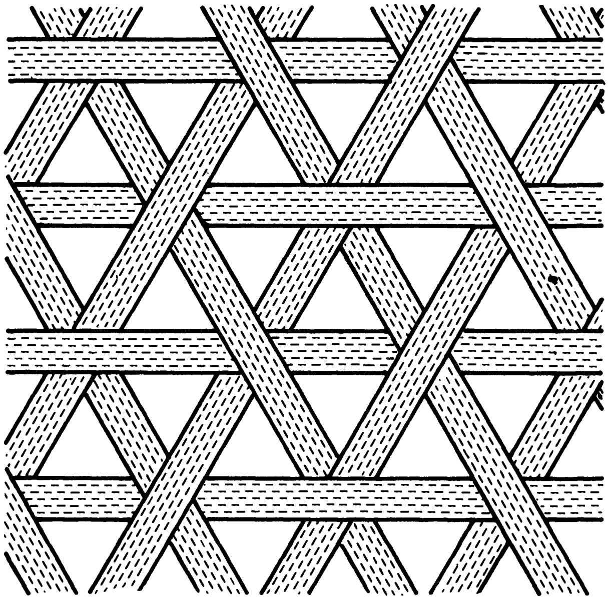


FIG. 16. A 3-way, 3-fold isonemal pseudofabric. Despite differences in overall appearance, it is not hard to verify that it is of the same henomeric type as the pseudofabric in FIGURE 15 and the "mad weave" fabric of FIGURE 3. The homeomeric types of these three weaves are all different.

weaves of FIGURES 4(c), 4(d), 15, and 16 are all isonemal; moreover, it is easy to verify that those shown in FIGURES 4(c), 15, and 16 are of the same henomeric type.

The main result of this article has an analogue for pseudofabrics. It can be proved that w-way, f-fold isonemal pseudofabrics exist if and only if $2 \leqslant f \leqslant w$ for $w = 2, 3, 4, 6$, or if (w, f) is one of the pairs $(2,4)$, $(3,4)$, $(3,5)$, $(3,6)$. In each case there are infinitely many fabrics of different henomeric types.

(f) Another generalization of fabrics, due to J. J. Pedersen, is to "fabrics" on polyhedral surfaces. A discussion of isonemal fabrics on such surfaces appears in [**36**].

(g) The first version of this article, entitled "The Geometry of Fabrics," was written in 1978 and presented at the Summer Meeting of the American Mathematical Society [**14**]; about a hundred copies were distributed at the meeting, or in response to requests. The publication was delayed, in part because small sections of that first version grew, on elaboration, into articles of considerable size; this led to publications [**15**], [**18**], [**19**], and [**20**]. The main result of the present paper (Theorem 5 of the 1978 version) has been described in several publications, in particular [**14**], [**17**], [**18**], and [**36**]; for related results concerning transitive families of straight lines see [**21**, Section 7.7]. Several other versions of the manuscript were prepared over the years, but at each stage there seemed to exist some reason for delaying the publication.

(h) The terminology of the various kinds of objects considered here has not been standardized so far; therefore, we have found it worthwhile to introduce different words for the different concepts of prefabrics, fabrics, and weaves (pseudofabrics). In all the cases we have tried to combine mathematical precision with common sense. The word "isonemal," introduced in [**14**], appears to have been universally accepted. Other terms (such as mononemal, twillin, w-way, f-fold, etc.) which we introduced in the various papers seem also to have met acceptance, or at least aquiescence. The word "fabric" has also been accepted quite generally. In some publications (see [**23**], [**24**]) the authors use "fabric" for what we call prefabrics. There is clearly no objection to authors using words in whatever manner they find convenient, but we strongly deplore the fact that they make no distinction between fabrics and prefabrics, *and even attempt to justify the fact that no such distinction is made*. (Our criticism in [**19**] that this gave rise to unnecessary confusion led to the polemic in [**25**], but the statements made there are fallacious besides misrepresenting the facts and the sources quoted.) We hope that the present paper will contribute to an accepted terminology.

(i) Several recent papers deal with 3-way fabrics. As already mentioned, [**4**] and [**9**] consider criteria for deciding whether a given 3-way prefabric hangs together. In [**34**] a study is begun of 3-way weaves which are the analogues of twills. In [**11**] several 3-way weaves are constructed, and the effects of using strands of various colors in the "mad weave" are investigated.

8. Practical uses of 3-way fabrics. Until very recently there seem to be no records or samples of *woven textiles* with strands in three or more directions; for information about the present situation see [**5**, p. 172], [**39**], [**41**]. But in related fields—basketry and plaiting in their different forms—the uses of several 3-way weaves are widespread, and go back to very ancient times. (More complex weaves are also quite common; for example, a 4-way weave is often used for seats and backs of cane chairs. However, since these weaves are not isonemal, we shall not discuss them here.)

The isonemal 3-way weave most frequently used in basketry is the pseudofabric shown in FIGURE 4(c). It has been called "hexagonal plaiting" or "hexagonal work" ([**31**], [**33**]), but these terms have also been applied to more complicated weaves. Its popularity is probably a result of its great strength and rigidity combined with light weight. For uses of this weave (and related ones) in various cultures, in objects ranging from cradles and crabpots to hats and snowshoes, see, for example [**7**], [**27**], [**28**], [**29**], [**35**], [**37**].

Another very popular 3-way isonemal weave (actually, a fabric) is the "mad weave." It is, apparently, quite old in various cultures of Southeast Asia. The first description of it in publications accessible to the West is by Bland [**1**] in 1906. The name of that weave seems to be related to the difficulties one may experience in practical work with it. These difficulties led to several detailed "how-to" descriptions [**12**, pp. 111-133], [**22**, pp. 172–183], [**32**], [**40**]. In fact, the "mad weave" as such is not all that hard—practical difficulties arise only if it is desired to enclose a

volume as for a basket, or to weave a polyhedral surface like one of those described in [**36**]. It appears that in the original version described in [**1**], the final product was a 3-way, 6-fold fabric! Moreover, although this 6-fold fabric is of the simple type, obtained by using a double thickness of each strand, the actual weaving was done the "hard way": the material used had one shiny side, and the first step consisted in producing a 3-fold weave with one shiny side; then the other three layers were worked into this weave in such a way that the other side also became shiny! This seems to be the only kind of isonemal 6-fold weave which has been found in objects of practical use.

REFERENCES

1. L. E. Bland, A few notes on the "Anyam Gila" basket making at Tanjong Kling, Malacca, Journal of the Straits Branch of the Royal Asiatic Society, 46(1906), pp. 1–7 and plates I–VI.
2. C. R. J. Clapham, When a fabric hangs together, Bull. London Math. Soc., 12(1980) 161–164.
3. ______, The bipartite tournament associated with a fabric, Discrete Math., 57(1985) 195–197.
4. ______, When a three-way fabric hangs together, J. Combinat. Theory, B38(1985) 190.
5. B. P. Corbman, Textiles: Fiber to Fabric, McGraw-Hill, New York, 1975.
6. H. S. W. Coxeter and W. O. J. Moser, Generators and Relations for Discrete Groups, fourth ed., Springer, Berlin, 1980.
7. S. C. Dellinger, Baby cradles of the Ozark Bluff Dwellers, American Antiquity, 1(1935/36) 197–214.
8. P. Engel, Über Wirkungsbereichsteilungen von kubischer Symmetrie, Zeitschrift für Kristallographie, 154(1981) 199–215.
9. T. C. Enns, An efficient algorithm determining when a fabric hangs together, Geometriae Dedicata, 15(1984) 259–260.
10. L. Fejes Tóth, Regular Figures, Pergamon, New York, 1964.
11. D. S. Fielker, Weaving tessellations, Mathematics Teaching, 91(1980) 34-39; 92 (1980) 40–43.
12. S. Glashausser and C. Westfall, Plaiting Step-By-Step, Watson-Guptill, New York, 1976.
13. Z. J. Grosicki, Watson's Advanced Textile Design, Newnes-Butterworths, London, 1977.
14. B. Grünbaum and G. C. Shephard, Geometry of Fabrics, Abstract *757-D1, Notices Amer. Math. Soc., 25(1978) A-462.
15. ______, Satins and twills: an introduction to the geometry of fabrics, Mathematics Magazine, 53(1980) 139–161.
16. ______, Tilings with congruent tiles, Bull. Amer. Math. Soc., N.S., 3(1980) 951–973.
17. ______, Tilings, patterns, fabrics and related topics in discrete geometry, Jahresberichte Deutsch. Math.-Verein., 85(1983) 1–32.
18. ______, The geometry of fabrics, in Geometrical Combinatorics, F. C. Holroyd and R. J. Wilson (eds.), Pitman, Boston-London-Melbourne, 1984, pp. 77–98.
19. ______, A catalogue of isonemal fabrics, in Discrete Geometry and Convexity, J. E. Goodman et al. (eds.), Ann. New York Acad. Sci., 440(1985) 279–298.
20. ______, An extension to the catalogue of isonemal fabrics, Discrete Math., 60(1986) 155–192.
21. ______, Tilings and Patterns, Freeman, New York, 1986.
22. V. E. Harvey, The Techniques of Basketry, Van Nostrand, New York-London, 1978.
23. J. A. Hoskins, W. D. Hoskins, A. P. Street, and R. G. Stanton, Some elementary isonemal binary matrices, Ars Combinatoria, 13(1982) 3–38.
24. J. A. Hoskins, R. G. Stanton, and A. P. Street, Enumerating the compound twillins, Congressus Numerantium, 38(1983) 3–22.
25. ______, Binary interlacement arrays, and how to find them, Congressus Numerantium, 42(1984) 321–376.

26. W. D. Hoskins and R. S. D. Thomas, Conditions for isonemal arrays on a Cartesian grid, Linear Algebra and its Appl., 57(1984) 87–103.
27. H. Krucker, Westafrikanische Mattengeflechte. Mitt. Ostschweiz., Geogr.-Kommerz. Gesell., St. Gallen, 1940/41.
28. K. Kudo and K. Suganuma, Japanese Bamboo Baskets, Kodansha International, Tokyo, 1980.
29. B. Laufer, Chinese Baskets, Anthropology Design Series No. 3, Field Museum of Natural History, Chicago, 1925.
30. G. E. Martin, Transformation Geometry: An Introduction to Symmetry, Springer-Verlag, New York-Heidelberg-Berlin, 1982.
31. O. T. Mason, Vocabulary of Malaysian basketwork: a study in the W. L. Abbott collections, Proc. U.S. National Museum (Washington, D.C.), 35(1908), pp. 1–51, plates 1–17.
32. ______, Anyam gila (mad weave): a Malaysian type of basket work, Proc. U.S. National Museum (Washington, D.C.), 36(1909) 385–390.
33. H. Miner, The importance of textiles in the archaeology of the Eastern United States, American Antiquity, 1(1935/36) 181–192.
34. R. Nowakowski and D. Skillicorn, Three-way weaving, Congressus Numerantium, 36(1982) 153–159.
35. L. Parker, Some common baskets of the Philippines, The Philippine Craftsman, 3(1914) 1–25.
36. J. J. Pedersen, Isonemal fabrics on polyhedral surfaces, in The Geometric Vein: The Coxeter Festschrift, C. Davis et al. (eds.), Springer-Verlag, New York-Heidelberg-Berlin, 1981, pp. 99–122.
37. G. Raval, A year's course in elementary bamboo weaving, The Philippine Craftsman, 3(1914) 247–262.
38. D. Schattschneider, The plane symmetry groups: their recognition and notation, this MONTHLY, 85(1978) 439–450.
39. A. Seiler-Baldinger, Systematik der Textilen Techniken, Basler Beiträge zur Ethnologie, vol. 14, Pharos-Verlag, Basel, 1973.
40. R. E. Spencer, The hexagonal weave basket, The Philippine Craftsman, 3(1914) 420–428.
41. W. C. Trost, Triaxial weaving: a new dimension in fabric formation. Samples of triaxially-woven fabrics: new directions to superior fabric performance, Leaflets issued by the Barber-Colman Company, Textile Machinery Division, Rockford, IL, 1979.

Miscellanea

A Mathematical Poem That Rhymes in London but Not in New York.

"It's elementary," the math prof said.
"Algebra's as simples as x, y, z."

Neal Hart
Department of Mathematics
Sam Houston State University
Huntsville, TX 77341

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

A Dozen Difficult Diophantine Dilemmas

ANDREW BREMNER
Department of Mathematics, Arizona State University, Tempe, AZ 85287

RICHARD K. GUY
Department of Mathematics & Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4

Most diophantine equations present a dilemma. Do you reach for a computer or a cold towel? Do you try to prove nonexistence of solutions or do you try to find the general solution, or at least an infinite family of solutions? Diophantine equations illustrate the dilemma of mathematical research in a very pure form: Do you try to prove a theorem, or do you search for a counterexample?

E. T. Prothro (1120 S. Duck, Stillwater, OK 74074) asks us to show that the equation

$$xy(x^4 - y^4) = 2zw(z^4 - w^4) \tag{1}$$

has no nontrivial solution in integers, or to exhibit such a solution.

By coincidence, in *Crux Mathematicorum* 10 (1984) Problem 994, Ernest J. Eckert asked if there are two distinct primitive Pythagorean triples whose products are equal, i.e., a nontrivial solution of

$$xy(x^4 - y^4) = zw(z^4 - w^4). \tag{2}$$

Computer searches over a small range failed to find any solution to (1) or (2). One feels instinctively that it is a difficult problem to decide whether there do indeed exist solutions. In geometric language, each of the equations (1) and (2) represents a sextic surface in three-dimensional projective space. There is reasonable knowledge of the arithmetic-geometric properties of surfaces of degrees 2, 3, and 4; but for higher degrees relatively little can be said. It is true, for instance, that on a nonsingular surface of degree greater than 4, there can exist only finitely many curves of genus zero.

It seems necessary to treat each such equation on its own merits, though we cannot think of any equation representing such a surface that has been satisfactorily solved by nontrivial methods. Perhaps the simplest example of the family is

$$x^6 + y^6 = z^6 + w^6 \tag{3}$$

for which, again, a complete solution is not known. By analogy with the recent proof by Faltings [**11**] of the Mordell conjecture, it may be that where some geometric invariants of a surface pass beyond a certain bound, there can be at most finitely many rational points on that surface.

Similarly, no nontrivial solution of

$$x^5 + y^5 = z^5 + w^5 \tag{4}$$

is known [D1 in **12**]. The paper [**14**] of Lander, Parkin, and Selfridge provides many other examples concerning sums of like powers: see also [**2**], [**3**], [**6**], [**7**].

No note on difficult diophantine equations is complete without a reference to Fermat's equation [D2 in **12**],

$$x^n + y^n = z^n, \qquad n \geqslant 3 \tag{5}$$

and to the equation*

$$x^4 + y^4 + z^4 = w^4 \tag{6}$$

associated with a conjecture of Euler [D1 in **12**].

A host of problems arises from Pythagorean triangles and their areas and medians. Dickson [**9**, p. 208] states that it "remains in doubt that no Heron triangle has more than one rational median." A Heron triangle is one with rational sides and area. Notice that the rationality of the area is what makes this problem difficult: for example, Euler noted that the triangle with sides 136, 174, and 170 has medians 158, 127, 131 (but area $240\sqrt{2002}$). We have *four* simultaneous equations, three applications of Apollonius's theorem and one of Heron's formula:

$$\begin{aligned} 2b^2 + 2c^2 &= a^2 + 4x^2 \\ 2c^2 + 2a^2 &= b^2 + 4y^2 \\ 2a^2 + 2b^2 &= c^2 + 4z^2 \\ 2b^2c^2 + 2c^2a^2 + 2a^2b^2 &= a^4 + b^4 + c^4 + 16\Delta^2. \end{aligned} \tag{7}$$

Recently, Ralph Heiner Buchholz has dispelled Dickson's doubt by finding a triangle with sides 146, 102, 52, area 1680 and medians 35, 97 and $4\sqrt{949}$, i.e., a Heron triangle (ABC in FIG. 1) with *two* rational medians. This implies the existence of a triangle (AMG in FIG. 1) with medians 73, 51, 26, area 560 and two rational sides. Randall L. Rathbun also found the triangle ABC, and another Heron triangle with sides 1750, 1252, 582, area 221760, and two rational medians, 1144 and 433. So the status of the problem is very similar to that of the rational cuboid [**15**]: we have seven quantities (here 3 sides, 3 medians, and the area; for the cuboid, 3 edge-lengths, 3 face-diagonals, and the body-diagonal) of which any six may be rational, but no one knows if all seven can be.

A further geometrically inspired problem is to find seven points, no four on a circle, no three in line, that define $\binom{7}{2}$ rational distances,

$$(x_i - x_j)^2 + (y_i - y_j)^2 = d_{ij}^2 \qquad (1 \leqslant i < j \leqslant 7). \tag{8}$$

*Noam Elkies, of Harvard University, has found an infinity of solutions of equation (6).

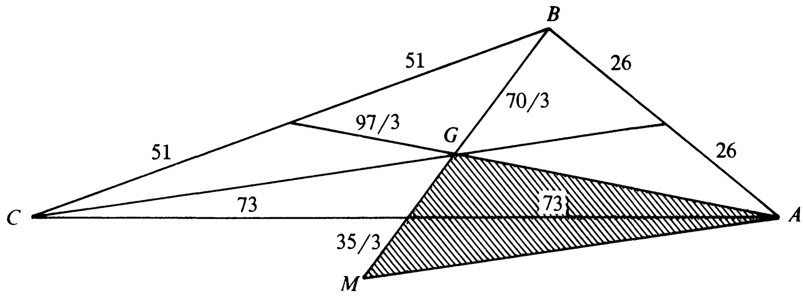


FIG. 1. Triangle whose sides are two-thirds of the medians of another triangle.

Here, only the d_{ij} are required to be rational, not necessarily the x_i, y_i.

It seems particularly difficult to determine if there is a point at integer distances a, b, c, d from the corners of a square of integer side s. Find solutions to the pair of simultaneous diophantine equations [D19 in **12**]:

$$\begin{gathered}(s^2+b^2-a^2)^2+(s^2+b^2-c^2)^2=4b^2s^2\\ a^2+c^2=b^2+d^2\end{gathered} \tag{9}$$

or prove that there aren't any. It is known [**5**], [**10**] how to construct infinitely many parametric solutions to the first of these, but no solution has yet been found which also satisfies the second.

Find all points at integer distances from the vertices of an integer sided equilateral triangle. If s is the side of the triangle, and x, y, z the distances, then the cosine formula yields

$$s^4+x^4+y^4+z^4=s^2x^2+s^2y^2+s^2z^2+y^2z^2+z^2x^2+x^2y^2 \tag{10}$$

and symmetry shows that solutions come in sets of 24. It is not hard, using Ptolemy's theorem [**8**, p. 42], to find an infinite number of solutions with the points on the sides, or on the circumcircle, of the equilateral triangle (FIG. 2).

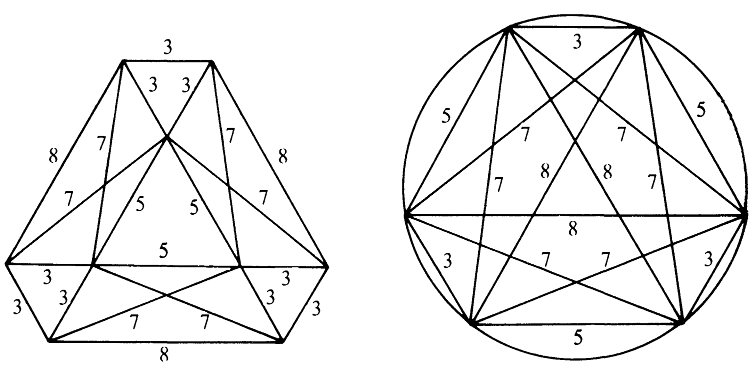


FIG. 2. Points at integer distances from the corners of an equilateral triangle.

However, Arnfried Kemnitz notes that $x = m^2 + n^2$; $y, z = m^2 \pm mn + n^2$ with $m = 2(u^2 - v^2)$, $n = u^2 + 4uv + v^2$ gives $s = 8(u^2 - v^2)(u^2 + uv + v^2)$ and an infinity of solutions in the which the point lies in general position with regard to the triangle. A computer search showed that (57, 65, 73, 112) was the smallest such solution (FIG. 3).

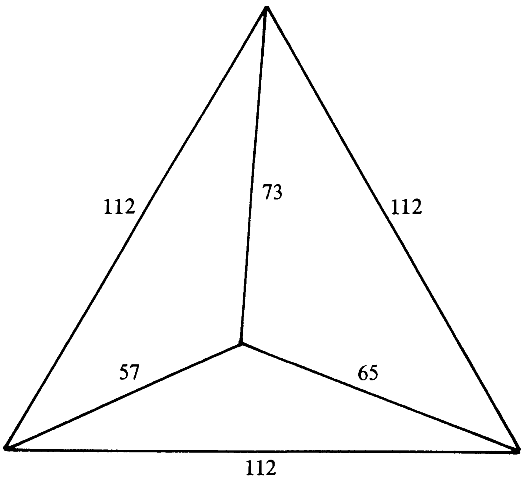


FIG. 3

Since we produced an earlier draft of this article, John Leech has sent us a neat and elementary proof of the fact that the points at rational distances from the vertices of *any* triangle with rational sides are dense in the plane of the triangle. This result was earlier proved by Almering [**1**].

Leech also observed that this problem has a sort of dual: find scalene triangles with integer sides, whose Fermat point (where the sides each subtend an angle of 120°) is at integer distances from the vertices. This requires that each of

$$y^2 + yz + z^2 \quad z^2 + zx + x^2 \quad x^2 + xy + y^2$$

be made a square. Dickson [**9**, p. 511] indicates that J. Cunliffe did this in 1809 by equating the first two expressions to $(y + z - m)^2$, $(z + x - n)^2$, giving x and y in terms of z (and m and n), whereupon the third expression gives a quartic in z, which is "made square in a classical manner." There is a strong similarity to the rational cuboid problem [**15**], [**17**]. We are dealing with the intersection of three quadrics in five-dimensional projective space. Such an object is a surface (in general, of type K3), but here, as in the case of the rational cuboid, the surface possesses singular points. At present, the general solution of such equations seems to be beyond reach.

Leech also asked (85-05-20) for triads of squares with equal sum and equal product:

$$x^2 + y^2 + z^2 = a^2 + b^2 + c^2, \qquad x^2y^2z^2 = a^2b^2c^2. \tag{11}$$

For example, $3^2, 8^2, 26^2$ and $2^2, 13^2, 24^2$. He writes, "I have a few others, but no systematic method."

Mark Samuelovitch Weissmann writes from Minsk to observe that $(1, -1, a+1, -a)$ gives an infinity of solutions of Mordell's equation [D28 in **12**],

$$\frac{1}{x} + \frac{1}{y} + \frac{1}{z} + \frac{1}{w} + \frac{1}{wxyz} = 0 \tag{12}$$

and that there are others, e.g., $(-2, 3, 7, 41)$. In fact, if we write $x = u - z$, $y = v - z$, $w = t - z$, then $2z^3 - z^2(t + u + v) + tuv + 1 = 0$. There are plenty of solutions, e.g., $x = 1 - z$; $y, w = 2z^2 - 2z \pm 1$, or, more generally, if $z^2 - z - 1 = \alpha\gamma$, $z^2 - z + 1 = \beta\delta$, $y = z(z-1) + \alpha\beta$, $w = z(z-1) + \gamma\delta$. For example, each of $z = 173$, 488, 668, and 983 gives 128 solutions.

As a warning not to reach too rapidly for your computer, glance at the paper [**4**]. A more recently worked example in similar vein is that the smallest rational solution of the equation $y^2 = x(x^2 + 1877)$ is

$$x = \frac{488\,114890\,132279\,420338\,923511\,111187\,227812\,402249\,922481}{1170\,488036\,631991\,182951\,963428\,713037\,871725\,514476\,328100}$$

$$y = \frac{1\,120418\,474814\,606797\,599253\,948625\,207508\,935768\,951951\,287850\,506482\,622440\,320026\,076879}{40045\,200163\,862032\,520789\,539685\,815389\,316574\,590961\,010977\,921502\,391797\,012063\,929000}.$$

Clearly one sometimes needs to think before rushing to the computer!

Send single solutions to any of equations (1) to (9), or general solutions to equations (10), (11), and (12).

REFERENCES

1. J. H. J. Almering, Rational quadrilaterals, *Indagationes Math.*, 25 (1963) 192–199; MR 26 #4963.
2. Andrew Bremner, A geometric approach to equal sums of sixth powers, *Proc. London Math. Soc.*, (3) 43 (1981) 544–581; MR 83g:14018.
3. ______, A geometric approach to equal sums of fifth powers, *J. Number Theory*, 13 (1981) 337–354; MR 83g:14017.
4. Andrew Bremner and J. W. S. Cassels, On the equation $Y^2 = X(X^2 + p)$, *Math. Comput.*, 42 (1984) 257–264; MR 85f:11017.
5. Andrew Bremner and Richard K. Guy, The delta-lambda configurations in tiling the square, *J. Number Theory* (to appear).
6. S. Brudno, Some new results on equal sums of like powers, *Math. Comput.*, 23 (1969) 877–880; MR 41#1646.
7. ______, On generating infinitely many solutions of the diophantine equation $A^6 + B^6 + C^6 = D^6 + E^6 + F^6$, *Math. Comput.*, 24 (1970) 453–454; MR 42#5903.
8. H. S. M. Coxeter and S. L. Greitzer, Geometry Revisited, New Math. Library 19, Random House, 1967.
9. L. E. Dickson, History of the Theory of Numbers, vol. II, G. E. Stechert & Co., New York, 1934.
10. R. B. Eggleton, A. S. Fraenkel, Richard K. Guy and J. L. Selfridge, Tiling the square with rational triangles, *Acta Arith.* (submitted).

11. Gerd Faltings, Endlichkeitssätze für abelsche Varietäten über Zahlkörpen, *Invent. Math.*, 73 (1983) 349–366; erratum *ibid.*, 75 (1984) 381; MR 85g:11026.
12. Richard K. Guy, Unsolved Problems in Number Theory, Springer, 1981.
13. Heiko Harborth and Arnfried Kemnitz, Diameters of integral point sets, *Colloq. Math. Soc. János Bolyai*, 48, Intuitive Geometry, Siófok, 1985.
14. L. J. Lander, T. R. Parkin, and J. L. Selfridge, A survey of equal sums of like powers, *Math. Comput.*, 21 (1967) 446–459; MR 36#5060.
15. J. Leech, The rational cuboid revisited, this MONTHLY, 84 (1977) 518–533; MR 56#5421; corrections, 85 (1978) 473; MR 58#16492.
16. L. J. Mordell, Diophantine Equations, Academic Press, 1969.
17. A. Bremner, The rational cuboid and a quartic surface, *Rocky Mountain J. Math.* (to appear).

Miscellanea

To be wholly devoted to some intellectual exercise is to have succeeded in life; and perhaps only in law and the higher mathematics may this devotion be maintained, suffice to itself without reaction, and find continual rewards without excitement.

from *Weir of Hermiston*, by Robert Louis Stevenson

NOTES

EDITED BY DENNIS DETURCK, RICHARD LIBERA, AND ANITA E. SOLOW

A Characterization of Graphs of Diameter Two

GARY S. BLOOM[1]
Computer Science Department, City College, CUNY, New York, NY 10031

JOHN W. KENNEDY[2] AND LOUIS V. QUINTAS[3]
Mathematics Department, Pace University, New York, NY 10038

In a recent Note [4] and a Letter to the Editor [5] the relationship between the diameter of a graph and the diameter of its complement was considered. The authors of [4] noted that the simplicity of their method of establishing the Ringel-Sachs relation was tantamount to including it in an ultimately understandable, apocalyptic "Divine Book of Proofs."

We propose another related entry in the awaited tome that shows clearly how graphs of diameter two can be characterized by a simple condition on their complements. This simple condition, in turn, implies several computationally convenient conditions which are sufficient to show that a graph has diameter two. When considered in appropriate fashion, these results are all easily grasped; nevertheless, we have found that they have considerable utility as lemmas to simplify potentially complex arguments (as exemplified in [1] and [2]). With the exception of our Theorem 2a, which is proved in [3] using five cases and subcases, we have not seen these results proved elsewhere. Parenthetically, we note that the theorem in [5] also appears as a lemma in [3].

The graph theory terminology we use is as follows. The *complete graph* K_p has p vertices with every two vertices adjacent (i.e., forming an edge). The *complement* $\bar{G}$ of a graph G is the graph with the same vertex set as G and precisely those edges that are not in G. A graph with no edges is called an *empty* graph. The *distance* between vertices x and y in G is the minimum number of edges possible in a path from x to y. The *diameter* $\operatorname{diam}(G)$ is the maximum distance which occurs in G, with the convention that $\operatorname{diam}(G) = \infty$ when G is not connected.

Let x be a vertex in G. A *star centered at* x consists of any set of edges in G which have x as a vertex. Let uv be an edge in G. A (maximal) *double-star on* uv is a maximal tree in G which is the union of stars centered at u or v such that each star contains the edge uv. A subgraph H of G is said to *span* G if H has the same vertex set as G.

[1] Research partially supported by PSC-CUNY Research Award 6-63230 of the City University of New York.

[2,3] Research partially supported by Pace University Scholarly Research Committee and Dyson College Faculty Research Grants.

THEOREM 1. *A graph G has diameter two if and only if $\overline{G}$ is not empty and $\overline{G}$ is not spanned by a double-star.*

Proof. $\overline{G}$ is empty if and only if $G = K_p$ and has diameter 0 or 1.

If $\overline{G}$ is not spanned by a double-star, then in $\overline{G}$, for every edge uv there is a vertex x that is neither adjacent to u nor to v. Thus, in G, x is the intermediate vertex in a path of length 2 between u and v. On the other hand, if u and v are not adjacent in $\overline{G}$, then the edge uv is present in G. Consequently, the distance between any pair of vertices in G is no more than 2.

If $\overline{G}$ is spanned by a double-star on edge uv, then in $\overline{G}$ every vertex x is adjacent to at least one of the points u or v. Thus, in G, every path between u and v must have length at least 3. □

A vertex x has *degree* r if x is the center of a maximal star with r edges. A graph is *r-regular* if each vertex of the graph has degree r.

THEOREM 2. *A graph G on p vertices has* $\operatorname{diam}(G) = 2$, *if* $G \neq K_p$ *and any one of the following conditions holds*:

(a) $\operatorname{diam}(\overline{G}) \geqslant 4$.

(b) $\operatorname{diam}(\overline{G}) = 3$ *and $\overline{G}$ is r-regular.*

(c) $p > d_1 + d_2$; *where* $d_1 \geqslant d_2 \geqslant \cdots \geqslant d_p$ *with d_i the degree of vertex i in $\overline{G}$.*

Two observations are worth noting here: When $\overline{G}$ is r-regular, condition (c) reduces to "$p > 2r$." Secondly, when $\operatorname{diam}(\overline{G}) = 2$, only condition (c) of Theorem 2 can apply. These observations can be used to infer, for example, that the diameter of the complement of the well-known Petersen graph is two, since the Petersen graph, which has diameter two, is 3-regular and has $p = 10$.

Proof. (a) Use Theorem 1 to observe that $\operatorname{diam}(\overline{G}) \geqslant 4$ implies that there is no double-star which spans $\overline{G}$.

(b) Pick vertices x and y that are at distance 3 in $\overline{G}$. Thus, stars centered at x and at y are disjoint. Since $\overline{G}$ is r-regular, the number of vertices in each of the maximal stars centered at x and y is $r + 1$. Thus, $\overline{G}$ has at least $2r + 2$ vertices. Since a double-star in $\overline{G}$ has no greater than $2r$ vertices, a double-star cannot span $\overline{G}$.

(c) Every double-star in $\overline{G}$ has at most $d_1 + d_2$ vertices. Thus, $p > d_1 + d_2$ implies that $\overline{G}$ is not spanned by any double-star. □

REFERENCES

1. D. Bauer, G. S. Bloom, and F. Boesch, Edge to Point Degree Lists and Triangle Free Point Degree Regular Graphs, pp. 93–104 in Progress in Graph Theory, J. A. Bondy and U. S. R. Murty (eds.), Academic Press, Toronto, 1984.
2. G. S. Bloom, J. W. Kennedy, and L. V. Quintas, Distance Degree Regular Graphs, pp. 95–108 in Theory and Applications of Graphs, G. Chartrand et al. (eds.), John Wiley & Sons, New York, 1981.
3. J. Bosák, A. Rosa, and Š. Znám, On Decompositions of Complete Graphs into Factors With Given Diameters, pp. 37–56 in Theory of Graphs, P. Erdös and G. Katona (eds.), Academic Press, New York, 1968.
4. F. Harary and R. W. Robinson, The Diameter of a Graph and its Complement, this MONTHLY, 92 (1985) 211–212.
5. P. D. Straffin, Jr., Letter to the Editor, this MONTHLY, 93 (1986) 76.

The Jordan-Brouwer Separation Theorem for Smooth Hypersurfaces

ELON L. LIMA
I.M.P.A., Estrada Dona Castorina 110, 22460 Rio de Janeiro, Brazil

We give here a simple proof of the following:

JORDAN-BROUWER SEPARATION THEOREM. *Let $M \subset \mathbb{R}^m$ be a connected, compact, orientable smooth hypersurface. Its complement $\mathbb{R}^m - M$ has two connected components, each of which has M as its point set boundary.*

A subset $M \subset \mathbb{R}^m$ is called a *smooth hypersurface* when every point $x \in M$ belongs to an open set U, on which is defined a smooth function $\varphi\colon U \to \mathbb{R}$ with the following properties: i) $\operatorname{grad}\varphi(x) \neq 0$; ii) $\varphi^{-1}(0) = M \cap U$. A vector $v \in \mathbb{R}^m$ is said to be *normal* to M at x when v is a multiple of $\operatorname{grad}\varphi(x)$. The *tangent space* to M at x is the set T_xM of all vectors in $\mathbb{R}^m$ that are perpendicular to $\operatorname{grad}\varphi(x)$. A map $f\colon M \to \mathbb{R}^n$ is called *smooth* when, for every $x \in M$, f is the restriction to $U \cap M$ of a smooth map $F\colon U \to \mathbb{R}^n$, defined on an open set U containing x. The *derivative* of a smooth map $f\colon M \to \mathbb{R}^n$ is the linear map $f'(x)\colon T_xM \to \mathbb{R}^n$, obtained by restriction of $F'(x)\colon \mathbb{R}^m \to \mathbb{R}^n$. (Recall that the matrix of $F'(x)$ is the Jacobian matrix of F.) A *diffeomorphism* is a smooth map with a smooth inverse. The Inverse Mapping Theorem says that if $f'(x)\colon T_xM \to \mathbb{R}^{m-1}$ is a linear isomorphism then f, restricted to some neighborhood V of x in M, gives a diffeomorphism of V onto an open subset of $\mathbb{R}^{m-1}$. (For more details, see Thorpe [**2**].)

A smooth hypersurface $M \subset \mathbb{R}^m$ is said to be *orientable* when it admits a smooth field of normal unit vectors, i.e., when there exists a smooth map $v\colon M \to \mathbb{R}^m$ such that $|v(x)| = 1$ and $v(x)$ is normal to M at x, for every $x \in M$.

The assumption of orientability in Theorem A is redundant: any compact hypersurface must be orientable. (See Samelson [**1**] for a short proof of the smooth case.) Its presence, however, makes possible an easy proof. In many cases (for instance, when $M = S^{m-1}$) orientability is known *a priori*.

By "smooth" we mean C^∞. The proof holds verbatim for C^2 surfaces and, with a small technical modification (transverse, instead of normal fields) it would apply for C^1 surfaces as well.

Samelson's method of proof may also be used to get Theorem A, but we believe that our approach is more elementary. Instead of the Transversality Theorem and the classification of one-dimensional manifolds, we use the well-known fact that any smooth vector field $X\colon \mathbb{R}^m \to \mathbb{R}^m$, $X(x) = (a_1(x), \ldots, a_m(x))$, which fulfills the *integrability conditions* $\partial a_i/\partial x_j = \partial a_j/\partial x_i$ $(i, j = 1, \ldots, m)$, is the gradient of a smooth function $\varphi\colon \mathbb{R}^m \to \mathbb{R}$.

Once and for all, we fix a smooth field of unit normal vectors $v\colon M \to \mathbb{R}^m$ and define a smooth map $h\colon M \times \mathbb{R} \to \mathbb{R}^m$ by $h(x, t) = x + t \cdot v(x)$. For any given $\varepsilon > 0$, we denote by $h_\varepsilon\colon M \times (-\varepsilon, \varepsilon) \to \mathbb{R}^m$ the restriction of h.

The following standard result is included here for completeness.

LEMMA. *Let $M \subset \mathbb{R}^m$ be a compact, orientable, smooth hypersurface. For some $\varepsilon > 0$, h_ε: $M \times (-\varepsilon, \varepsilon) \to \mathbb{R}^m$ is a diffeomorphism onto an open subset of $\mathbb{R}^m$.*

Proof. For every $x \in M$, the derivative $h'(x,0)$: $T_xM \times \mathbb{R} \to \mathbb{R}^m$ is a linear isomorphism, since it sends each horizontal vector $(w,0)$ into w and each vertical vector $(0,t)$ into $t \cdot v(x)$ (which is perpendicular to w). By the Inverse Mapping Theorem, we may find $\delta_x > 0$ and an open neighborhood V_x of x in M such that h maps $V_x \times (-\delta_x, \delta_x)$ diffeomorphically onto an open neighborhood of x in $\mathbb{R}^m$. It remains only to show that, for some $\varepsilon > 0$, h_ε is injective. Assuming otherwise, we would find, for each $n \in \mathbb{N}$, *distinct* pairs $(x_n, s_n), (y_n, t_n)$ in $M \times (-1/n, 1/n)$ such that $h(x_n, s_n) = h(y_n, t_n)$. Since $M \times [-1,1]$ is compact we may assume (by taking subsequences, if necessary) that $x_n \to x \in M$, $y_n \to y \in M$, $s_n \to 0$ and $t_n \to 0$. Then

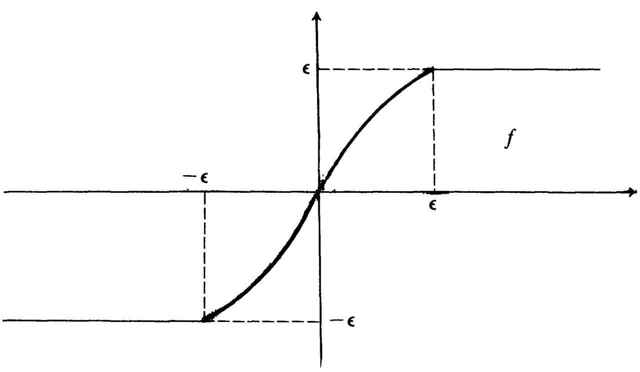
$$x = h(x,0) = \lim_n h(x_n, s_n) = \lim_n h(y_n, t_n) = h(y,0) = y.$$

Hence $\lim(x_n, s_n) = \lim(y_n, t_n) = (x,0)$. For all large values of n, (x_n, s_n) and (y_n, t_n) would belong to $V_x \times (-\delta_x, \delta_x)$ and then $h(x_n, s_n) \neq h(y_n, t_n)$. This contradiction proves the lemma.

We denote the image of h_ε by $V_\varepsilon(M)$ and call it a *tubular neighborhood of M*. We also write $V_\varepsilon[M] = h(M \times [-\varepsilon, \varepsilon])$.

Proof of the Theorem. We begin by showing that $\mathbb{R}^m - M$ is disconnected. More precisely, we define a smooth function φ: $\mathbb{R}^m \to \mathbb{R}$ such that $M = \varphi^{-1}(0)$ and grad $\varphi(x) \neq 0$ for every $x \in M$. Then the open sets $A = \{x \in \mathbb{R}^m; \varphi(x) > 0\}$ and $B = \{x \in \mathbb{R}^m; \varphi(x) < 0\}$ are nonempty, disjoint, with $\mathbb{R}^m - M = A \cup B$. The definition of φ is as follows.

Let $V_{2\varepsilon}(M)$ be a tubular neighborhood of M. Take a smooth function f: $\mathbb{R} \to \mathbb{R}$ such that $f(0) = 0$, $f'(t) > 0$ when $-\varepsilon < t < \varepsilon$, $f(t) = \varepsilon$ for $t \geqslant \varepsilon$ and $f(t) = -\varepsilon$ for $t \leqslant -\varepsilon$.



Any point in $V_{2\varepsilon}(M)$ may be written uniquely as $x + t \cdot v(x)$, with $x \in M$ and $|t| < 2\varepsilon$. Define a function g: $V_{2\varepsilon}(M) \to \mathbb{R}$ by $g(x + t \cdot v(x)) = f(t)$. Then g is

smooth, $M = g^{-1}(0)$, $g(x + t \cdot v(x)) = \varepsilon$ for $t \in [\varepsilon, 2\varepsilon)$ and $g(x + t \cdot v(x)) = -\varepsilon$ for $t \in (-2\varepsilon, -\varepsilon]$. Let X: $\mathbb{R}^m \to \mathbb{R}^m$ be the vector field which equals gradφg on $V_{2\varepsilon}(M)$ and vanishes outside this tubular neighborhood. The set $V_\varepsilon[M]$ is compact, hence closed in $\mathbb{R}^m$. X is smooth on each of the open sets $V_{2\varepsilon}(M)$ and $\mathbb{R}^m - V_\varepsilon[M]$ (in fact, identically zero on the latter). So, X is a smooth vector field in $\mathbb{R}^m$, which clearly fulfills the integrability conditions. Hence we may find a smooth function φ: $\mathbb{R}^m \to \mathbb{R}$ such that grad$\varphi = X$.

By adding a constant to φ, if necessary, we may assume that $\varphi(x) = g(x)$ for every x in $V_{2\varepsilon}(M)$, since this is a connected open set on which φ and g have the same gradient. Moreover, φ is constant on every connected component of $\mathbb{R}^m - V_\varepsilon(M)$, since its gradient vanishes there. Now, every such component meets $V_{2\varepsilon}(M)$. [Given any $y \notin V_{2\varepsilon}(M)$, let p be a point in the closed set $V_\varepsilon[M]$ that minimizes distance from y. The line segment $[y, p]$ lies in $\mathbb{R}^m - V_\varepsilon(M)$ (hence in the component of y in this set) and meets $V_{2\varepsilon}(M)$ at all points near p.] Then $\varphi = \pm\varepsilon$ outside $V_\varepsilon(M)$ and $M = \varphi^{-1}(0)$. If $x \in M$ then grad$\varphi(x) = c \cdot v(x)$, $c \in \mathbb{R}$, because the gradient is perpendicular to the level surface M. This gives

$$c = \langle \operatorname{grad} \varphi(x), v(x) \rangle = \frac{d}{dt}[g(x + t \cdot v(x))]_{t=0} = f'(0) = 0,$$

hence grad $\cdot\, \varphi(x) \neq 0$ at each point $x \in M$.

Next, we show that the open sets A, B, defined at the beginning of the proof, are connected. In fact, A contains the connected set $P = h(M \times (0, 2\varepsilon)) = \{x + t \cdot v(x);\ x \in M,\ 0 < t < 2\varepsilon\}$. Moreover, every $y \in A$ is either in P or may be joined to a point $p \in P$ by a line segment $[y, p] \subset A$: just take p, (as before) as a point in the closed set $h(M \times [0, \varepsilon])$ that minimizes distance from y. A similar argument proves that B is connected.

Finally, we prove that M is the common point set boundary of both connected components A, B of $\mathbb{R}^m - M$.

For any $x \in M$, the point $x + t \cdot v(x)$ belongs to A when $0 < t < \varepsilon$ and to B when $-\varepsilon < t < 0$. Let $fr \cdot S$ denote the point set boundary of a set S. This shows that $x \in fr \cdot A \cap fr \cdot B$ for all $x \in M$. On the other hand, if $x \in fr \cdot A$ then $\varphi(x) \geqslant 0$ because $x \in \bar{A}$, but $\varphi(x) \leqslant 0$ because $x \notin A$, so $\varphi(x) = 0$. Therefore $fr \cdot A \subset M$, that is, $fr \cdot A = M$. Similarly, $fr \cdot B = M$ and the proof is finished.

REMARK. The same kind of argument applies when the hypersurface M, instead of compact, is assumed only to be a *closed* subset of $\mathbb{R}^m$. One needs only to change the Lemma, where instead of a constant $\varepsilon > 0$, a continuous positive function ε: $M \to \mathbb{R}$ must be found with the following property: if $x|y$ are in M then $x + s \cdot v(x)|y + t.v(y)$ for all $s \in (-\varepsilon(x), \varepsilon(x))$ and $t \in (-\varepsilon(y), \varepsilon(y))$. One may even go one step further and replace $\mathbb{R}^m$ by any simply-connected m-dimensional surface N containing the $(m-1)$-dimensional surface M as a closed subset. The same idea still applies, except that the construction of the tubular neighborhood $V_\varepsilon(M) \subset N$ is more subtle. (Instead of straight line segments one may use geodesics.)

REFERENCES

1. H. Samelson, Orientability of Hypersurfaces in $\mathbb{R}^n$, Proc. Am. Math. Soc., 22 (1969) pp. 301–302.
2. J. A. Thorpe, Elementary Topics in Differential Geometry, Springer Verlag, New York, 1979.
3. (Added in proof) B. Doubrovine, S. Novikov, A. Fomenko, Géométrie Contemporaine, Mir, Moscow, 1979. (Page 67, vol. 2 of this remarkable book contains a 10-line proof of the above result. That proof is wrong.)

The Königsberg Bridges—250 Years Later

PETER A. FOWLER
Department of Mathematics, California State University, Hayward, CA 94542

The year 1736 is generally accepted as marking the beginning of graph theory. In that year Euler published his article [2] on the Königsberg bridges problem and its generalizations, in which he demonstrated the necessity that every vertex have even degree. Remarkably, he dismissed the more difficult sufficiency argument as an "easy task... after a little thought." Not until 1873 was a sufficiency proof published [3]. Since that time there have been many proofs published. (See [1] for translations of [2] and [3] and for an interesting history of the problem.) The usual approach considers an Euler tour as a circuit to be traced out dynamically, ever enlarging until the graph is covered. In honor of the 250th anniversary of Euler's paper, we give a new proof in which we view the Euler tour as a static object in a graph and employ induction.

THEOREM. *A connected multigraph in which each vertex is of even degree has an Euler tour, that is, a circuit containing all the edges of the graph.*

Proof. We proceed by induction on the number of edges. If a graph has fewer than three vertices, the result is true by inspection, so consider a graph G with at least three vertices and n edges. Clearly, n is at least three.

Assume that an Euler tour exists in any connected graph with all even degree vertices and fewer than n edges. Select any vertex v of G. Since G is connected and all vertices have even degree, there exist two distinct edges which join v to (not necessarily distinct) vertices u and w. Delete these two edges. If $u = w$, insert a loop at u; otherwise, join u to w by an edge. The resulting graph has all vertices of even degree and has $n - 1$ edges. If it is connected, it follows from the inductive hypothesis that it has an Euler tour. An Euler tour for G can be obtained by replacing u, uw, w by u, uv, v, vw, w. If the new graph is disconnected, then it has exactly two components one of which contains v and the other of which contains u

and w. Both components satisfy the conditions and have fewer than n edges, hence they contain Euler tours, say, $T1$ and $T2$. An Euler tour for G can be constructed by replacing u, uw, w in $T1$ by u, uv, $T2$, vw, w.

It is worth remarking that essentially the same proof can be given by using induction on the number of vertices.

The author thanks David Gale and Kenneth Rebman for their comments.

REFERENCES

1. N. L. Biggs, E. K. Lloyd, and R. J. Wilson, Graph Theory 1736–1936, Oxford University Press, Oxford, 1977.
2. L. Euler, Solutio Problematis ad Geometriam Situs Pertinentis, Commentarii Academiae Scientiarum Imperialis Petropolitanae, 8 (1736) 128–140.
3. C. Hierholzer, Uber die Möglichkeit, einen Linienzug ohne Wiederholung und ohne Unterbrechung zu umfahren, Math. Ann., 6 (1973) 30–32.

THE TEACHING OF MATHEMATICS

EDITED BY JOAN P. HUTCHINSON AND STAN WAGON

We were deeply saddened to learn of the death of Tory Parsons on April 2, 1987. His enthusiasm and energy for mathematics made it a pleasure and a privilege to have worked with him.

The Editors

ON MONOTONE SUBSEQUENCES

DONALD J. NEWMAN
Department of Mathematics, Temple University, Philadelphia, PA 19122

T. D. PARSONS*
Department of Mathematics, California State University, Chico, CA 95929

We will give a very short proof of the well-known fact that every infinite sequence $\{a_n\}$ of real numbers contains an infinite monotone subsequence. We will then comment on how this can best be used to advantage in teaching elementary courses in real analysis. Lastly, we will discuss connections between the proof and Ramsey's Theorem.

Let $S = \{i \mid \text{for all } j > i,\ a_i < a_j\}$. If S is infinite, with elements $i_1 < i_2 < \cdots$, then $a_{i_1} < a_{i_2} < \cdots$ is monotone strictly increasing. Otherwise, S is finite. In that case, there is a least index i_0 such that for all $i \geq i_0$ we have $i \notin S$. Since $i_0 \notin S$, there exists $i_1 > i_0$ such that $a_{i_0} \geq a_{i_1}$. And $i_1 \notin S$ so there exists $i_2 > i_1$ such that $a_{i_1} \geq a_{i_2}$, and so forth. Then $a_{i_0} \geq a_{i_1} \geq a_{i_2} \geq \cdots$ is monotone nonincreasing, completing the proof.

Courses in elementary real analysis usually begin with the properties of the real number field, including the "least upper bound property" or one of its equivalents. It is then natural and transparently easy to prove that every bounded monotone sequence of real numbers converges to its least upper bound (for nondecreasing sequences) or its greatest lower bound (for nonincreasing sequences). The above proof then quickly implies the Bolzano-Weierstrass Theorem and the convergence of Cauchy sequences. The former follows because every bounded sequence contains a (bounded) monotone subsequence that must converge. The latter follows because every Cauchy sequence is bounded, and must converge to the same limit as any one of its (bounded) monotone subsequences.

Our proof of the existence of monotone subsequences appears not to be well known, though it has appeared in a problem book [2], and it was independently

*Supported in part by NSF grant DMS 8616041

rediscovered by Parsons while working on some problems related to Ramsey's Theorem. We have been unable to find it elsewhere, though we would expect that such a simple proof must have been discovered long ago. Our purpose in presenting it here is to publicize it to teachers of elementary analysis, and to point out its interesting connections with variants of Ramsey's Theorem.

The infinite version of Ramsey's Theorem states that in any coloring, with two colors, of the edges of the complete graph with vertices the positive integers, there must exist an infinite monochromatic complete subgraph. This implies the existence of a monotone subsequence in any sequence of reals: a sequence $\{a_n\}$ induces the coloring in which edge $\{i, j\}$, for $i < j$, is colored red if $a_i < a_j$ or blue if $a_i \geq a_j$. A monochromatic infinite complete subgraph with vertices $i_1 < i_2 < \cdots$ then yields the monotone subsequence $a_{i_1}, a_{i_2}, \ldots,$ which is increasing or nonincreasing according as the subgraph has its edges all red or all blue. (See [**1**, p. 17].)

There is a simple proof [**3**] of Ramsey's Theorem that is only slightly more complicated than our proof above for the existence of monotone subsequences. Our proof corresponds to a weaker type of Ramsey theorem, involving ordered sets. Namely, let K be the infinite directed complete graph with vertex-set the positive integers and with arcs the ordered pairs (i, j) such that $i < j$. Then every coloring (with colors red and blue) of the arcs of K contains either an infinite (directed) complete subgraph (isomorphic to K itself) with all its arcs colored red or else an infinite directed path with all its arcs blue.

The proof of this is essentially the same as that given in the second paragraph of this note: in the definition of the set S, just replace "$a_i < a_j$" by "arc (i, j) is colored red" and consider the cases where S is infinite or finite.

This Ramsey variant clearly implies the monotone subsequence theorem as a corollary. It also implies an undirected version in which the arrows on the edges are disregarded, so that a coloring then gives either an infinite red complete subgraph or an infinite blue path. And that version implies one in which we get an infinite monochromatic path. But these undirected versions are apparently each insufficient to prove the monotone subsequence theorem.

Finally, we note that our proof used only the property that the reals are a linearly ordered set. Therefore, any infinite sequence of elements from any linearly ordered set has a monotone subsequence.

We are grateful to Simon Goberstein, Neville Robbins, and the referee for some comments that improved this note.

REFERENCES

1. Ronald L. Graham, Bruce L. Rothschild, and Joel H. Spencer, Ramsey Theory, John Wiley and Sons, New York, 1980.
2. Donald J. Newman, A Problem Seminar, Springer-Verlag, New York, 1982 (problem 6, solution p. 45).
3. Joseph G. Rosenstein, Linear Orderings, Academic Press, New York, 1982 (pp. 111–112).

Why Does the Wronskian Work?

MARK KRUSEMEYER
Mathematics Department, Carleton College, Northfield, MN 55057

The approach to the question in the title that is presented here has become nonstandard, although it can be found in 19th century literature. This older approach is more easily motivated, and perhaps more convincing, than the "modern" approach.

Recall that if $f_1, \ldots, f_n$ are $(n-1)$ times differentiable functions of a real variable t, their *Wronskian* is the function $W(f_1, \ldots, f_n)$ defined by

$$W(f_1, \ldots, f_n)(t) = \begin{vmatrix} f_1(t) & f_2(t) & \cdots & f_n(t) \\ f_1'(t) & f_2'(t) & \cdots & f_n'(t) \\ \vdots & \vdots & & \vdots \\ f_1^{(n-1)}(t) & f_2^{(n-1)}(t) & \cdots & f_n^{(n-1)}(t) \end{vmatrix}.$$

The question in the title refers to the following standard result.

THEOREM 1. *If $f_1, \ldots, f_n$ are linearly dependent on some interval (with nonempty interior), then $W(f_1, \ldots, f_n)$ is identically zero on that interval. Conversely, if $W(f_1, \ldots, f_n) \equiv 0$ on an interval on which $f_1, \ldots, f_n$ are solutions to the homogeneous linear differential equation*

$$y^{(n)} + p_{n-1}(t)y^{(n-1)} + \cdots + p_1(t)y' + p_0(t)y = 0$$

with $p_0, \ldots, p_{n-1}$ defined and continuous, then $f_1, \ldots, f_n$ are linearly dependent on that interval.

Along with this result, it is customary to give a warning, as follows: If $W(f_1, \ldots, f_n) \equiv 0$ but nothing else is known about $f_1, \ldots, f_n$, it does *not* follow that $f_1, \ldots, f_n$ are linearly dependent.

This warning was first given by Peano **[8]**, **[9]**; he actually had to give it twice, because an editor added a mistaken footnote that contradicted the main point of Peano's first warning. One of Peano's examples, which is still often used, is $f_1(t) = t^2$, $f_2(t) = t|t|$, for which $W(f_1, f_2) \equiv 0$ but f_1, f_2 are linearly independent. (Note, however, that f_1, f_2 are linearly dependent if they are restricted either to the interval $(-\infty, 0]$ or to $[0, \infty)$.)

On the other hand, it seems unsatisfactory to have the second part of the theorem include a hypothesis involving the nth derivatives of $f_1, \ldots, f_n$ (since these don't occur in the Wronskian itself). Other hypotheses ensuring the linear dependence of

$f_1, \ldots, f_n$ when $W(f_1, \ldots, f_n) \equiv 0$ were studied by various authors after Peano, including Bôcher [1]. Bôcher gave several such criteria, including:

THEOREM 2. *Assume $W(f_1, \ldots, f_n) \equiv 0$ on an interval on which the Wronskian of the first $n-1$ functions does not vanish, i.e., $W(f_1, \ldots, f_{n-1})(t) \neq 0$ for all t. Then $f_1, \ldots, f_n$ are linearly dependent on that interval.*

By induction and using continuity, one quickly derives the following from this, as in Hurewicz [6, p. 48].

THEOREM 3. *If $W(f_1, \ldots, f_n) \equiv 0$ on an interval, then there exists a subinterval on which $f_1, \ldots, f_n$ are linearly dependent.*

Note that Theorem 3 implies the second part (which is the harder) of Theorem 1. For suppose $f_1, \ldots, f_n$ are linearly dependent on some subinterval, so we have $c_1 f_1 + c_2 f_2 + \cdots + c_n f_n \equiv 0$ there with not all the numbers $c_1, \ldots, c_n$ zero. Then in the situation of Theorem 1, both $c_1 f_1 + c_2 f_2 + \cdots + c_n f_n$ and 0 are solutions of the initial value problem

$$y^{(n)} + p_{n-1}(t) y^{(n-1)} + \cdots + p_1(t) y' + p_0(t) y = 0,$$
$$y(t_0) = 0,$$
$$y'(t_0) = 0, \ldots, y^{(n-1)}(t_0) = 0 \quad (\text{for any } t_0 \text{ in the subinterval}).$$

Therefore, by a well-known uniqueness theorem (for a proof see, e.g., Coddington [3, p. 259]), $c_1 f_1 + \cdots + c_n f_n$ must be identically 0 on the whole interval.

Although Hurewicz's approach clears up the situation considerably, it requires that Theorem 2 be proved first. There is, however, a direct proof of Theorem 3 using a recursive formula for the Wronskian, which formula was already known to Christoffel [2, p. 298] and Hesse [5, p. 249, 250]. The proof itself was apparently first published by Frobenius [4], who characterized it as "auf rein rechnendem Wege" ("purely computational"). Frobenius derived the recursion anew, without citing Christoffel or Hesse. He did not actually refer to the Wronskian (this term was coined by Muir in 1882; see [7]), but to the "determinant" of $f_1, \ldots, f_n$. Apparently he did not realize that his conclusion was only true for a subinterval, and some time after Peano's warning his method must have fallen into disuse.

The recursive formula for the Wronskian is as follows.

PROPOSITION. *$W(f_1, \ldots, f_n) = f_1^n W((f_2/f_1)', (f_3/f_1)', \ldots, (f_n/f_1)')$ wherever f_1 is not zero.*

This is proved using the

LEMMA. *For any $(n-1)$ times differentiable function g,*

$$W(f_1 g, f_2 g, \ldots, f_n g) = g^n W(f_1, \ldots, f_n).$$

Proof of the lemma.

$$W(f_1g, f_2g, \ldots, f_ng) = \begin{vmatrix} f_1g & f_2g & \cdots & f_ng \\ (f_1g)' & (f_2g)' & \cdots & (f_ng)' \\ \vdots & \vdots & & \vdots \\ (f_1g)^{(n-1)} & (f_2g)^{(n-1)} & \cdots & (f_ng)^{(n-1)} \end{vmatrix}$$

$$= \begin{vmatrix} f_1g & f_2g & \cdots & f_ng \\ f_1'g + f_1g' & f_2'g + f_2g' & \cdots & \\ f_1''g + 2f_1'g' + f_1g'' & \cdots & & \\ \vdots & & & \end{vmatrix}.$$

We can factor out g from the first row of this determinant, then subtract g' times the new first row from the second to obtain

$$W(f_1g, f_2g, \ldots, f_ng) = g\begin{vmatrix} f_1 & f_2 & \cdots & f_n \\ f_1'g & f_2'g & \cdots & f_n'g \\ f_1''g + 2f_1'g' + f_1g'' & \cdots & & \\ \vdots & & & \end{vmatrix}.$$

Now we can factor g from the second row and subtract g'' times the first row and $2g'$ times the new second row from the third. Continuing in this way, we obtain

$$W(f_1g, f_2g, \ldots, f_ng) = g^nW(f_1, \ldots, f_n),$$

as claimed.

Proof of the proposition. By the lemma, wherever f_1 is not zero,

$$W(f_1, \ldots, f_n) = f_1^nW(1, f_2/f_1, f_3/f_1, \ldots, f_n/f_1)$$

$$= f_1^n\begin{vmatrix} 1 & f_2/f_1 & f_3/f_1 & \cdots & f_n/f_1 \\ 0 & (f_2/f_1)' & (f_3/f_1)' & \cdots & (f_n/f_1)' \\ 0 & \vdots & & & \\ \vdots & & & & \\ 0 & (f_2/f_1)^{(n-1)} & \cdots & & \end{vmatrix}$$

$$= f_1^nW((f_2/f_1)', (f_3/f_1)', \ldots, (f_n/f_1)')$$

(by expanding along the first column).

Proof of Theorem 3. We use induction on n; the case $n = 1$ is clear. Assume $W(f_1, \ldots, f_n) \equiv 0$ on an interval. If $f_1 \equiv 0$ on the interval, we are done; otherwise, we can find a subinterval on which f_1 does not vanish. By the proposition,

$$W((f_2/f_1)', (f_3/f_1)', \ldots, (f_n/f_1)') \equiv 0$$

on this subinterval, so by induction we can find a subinterval on which $(f_2/f_1)', \ldots, (f_n/f_1)'$ are linearly dependent. Thus there are numbers $c_2, \ldots, c_n$, not all zero, for which

$$c_2(f_2/f_1)' + \cdots + c_n(f_n/f_1)' \equiv 0,$$

that is,

$$((c_2 f_2 + \cdots + c_n f_n)/f_1)' \equiv 0.$$

Therefore, $(c_2 f_2 + \cdots + c_n f_n)/f_1$ must be a constant c, and it follows that $f_1, \ldots, f_n$ are linearly dependent.

REFERENCES

1. M. Bôcher, Certain cases in which the vanishing of the Wronskian is a sufficient condition for linear dependence, *Transactions of the A.M.S.*, 2 (1901) 139–149.
2. E. B. Christoffel, Ueber die lineare Abhängigkeit von Functionen einer einzigen Veränderlichen, *J. Reine Angewandte Math.*, 55 (1858) 281–299.
3. E. A. Coddington, An Introduction to Ordinary Differential Equations, Prentice-Hall, 1961.
4. F. G. Frobenius, Über die Determinante mehrerer Functionen einer Variabelen, *J. Reine Angewandte Math.*, 77 (1874) 245–257; Gesammelte Abhandlungen I, 141–153.
5. O. Hesse, Über die Criterien des Maximums und Minimums der einfachen Integrale, *J. Reine Angewandte Math.*, 54 (1857) 227–273.
6. W. Hurewicz, Lectures on Ordinary Differential Equations, M.I.T.-Wiley, 1958.
7. T. Muir, The Theory of Determinants in the Historical Order of Development, Vol. II, Macmillan, 1911.
8. G. Peano, Sur le determinant Wronskien, *Mathesis*, 9 (1889) 75–76.
9. ______, Sur les Wronskiens, *Mathesis*, 9 (1889) 110–112.

PROBLEMS AND SOLUTIONS

EDITED BY PAUL T. BATEMAN, HAROLD G. DIAMOND, KENNETH B. STOLARSKY (ADVANCED PROBLEMS), AND DOUGLAS B. WEST (ELEMENTARY PROBLEMS)

*For instructions about submitting **proposed** problems for publication in this department see the inside front cover. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the address given on the inside front cover.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the begining. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

For instructions about submitting solutions of Problems, which should be mailed by May 31, 1988, see the inside front cover. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label if an acknowledgment is desired.

E 3243. *Proposed by William F. Trench, Trinity University, San Antonio, TX.*

Suppose that g is constant in sign and locally Riemann-integrable on $[0,+\infty)$. Suppose f is nonincreasing on $[0,+\infty)$ and $\int_0^\infty f(x)\,dx = +\infty$. Show that

$$\int_0^\infty |f(x)\cos x + g(x)\sin x|\,dx = \int_0^\infty |f(x)\sin x + g(x)\cos x|\,dx = +\infty.$$

E 3244. *Proposed by R. S. Rodriguez and H. Sherwood, University of Central Florida, Orlando.*

Suppose $a_0, a_1, \ldots, a_k$ are integers such that $0 < a_0 < a_1 < \cdots < a_k$. Put

$$f(x) = \frac{1}{(1+x)^{a_0}} + \frac{x}{(1+x)^{a_1}} + \frac{x^2}{(1+x)^{a_2}} + \cdots + \frac{x^k}{(1+x)^{a_k}}.$$

The function f is easily seen to be strictly decreasing on $(0,+\infty)$ when each $a_i = i+1$. Is f strictly decreasing for every permissible choice of $a_0, a_1, \ldots, a_k$?

E 3245. *Proposed by Bruce A. Reznick and Lee A. Rubel, University of Illinois at Urbana-Champaign.*

(a) For what positive integers n does there exist an n by n matrix A over $\mathbb{C}$ having the following three properties:

(i) $n^2 - n$ of the entries are zero,

(ii) there are n distinct nonzero entries $r_1, r_2, \ldots, r_n$, none of which lies on the main diagonal,

(iii) the eigenvalues of A are $r_1, r_2, \ldots, r_n$?

(b) Same as (a) except that all entries are required to be real.

E 3246. *Proposed by Solomon W. Golomb, University of Southern California.*

Define a function ψ on the positive integers by putting $\psi(n) = n\phi(n)$, where ϕ is Euler's function. (a) Is ψ an injective function? (b) For how many positive integers n is $\psi(n)$ a perfect square? (c) Given any positive integer m, prove that there is an integer n divisible by m such that $\psi(n)$ is a perfect cube.

E 3247. *Proposed by Nick M. Kinnon, Winchester College, Winchester, England.*

Given a positive integer k greater than 2, the k-gonal numbers are the terms of the sequence

$$1, k, 3k-3, 6k-8, \ldots,$$

which has constant second difference $k-2$. The Fermat numbers are the terms of the sequence

$$3 = 2^{2^0}+1, \quad 5 = 2^{2^1}+1, \quad 17 = 2^{2^2}+1, \quad 257 = 2^{2^3}+1, \ldots.$$

Under what circumstances can a k-gonal number be a Fermat number?

E 3248. *Proposed by Paul Erdős, Hungarian Academy of Sciences, and John Selfridge, Northern Illinois University.*

(a) Let S be the set of positive integers with no prime factors bigger than 3. Prove that every positive integer is expressible as a sum of distinct elements of S such that no summand is a multiple of any other. For example, $19 = 9 + 6 + 4$.

(b) Let T be the set of positive integers with no prime factors other than 2, 5, or 7. Prove that every sufficiently large positive integer is expressible as a sum of distinct elements of T such that no summand is a multiple of any other. For example, $62 = 49 + 8 + 5$.

SOLUTIONS OF ELEMENTARY PROBLEMS

Different Difference Bounds

E 3095 [1985, 428]. *Proposed by Alexandru Lupas, Facultatea de Mecanica, Sibiu, Romania.*

Let p be a fixed natural number, $(a_n)_{n=0}^{\infty}$ a sequence of real numbers, and

$$\Delta^p a_n = \sum_{k=0}^{p} (-1)^{p-k}\binom{p}{k} a_{n+k}, \qquad A_n = \frac{1}{n+1}\sum_{k=0}^{n} a_k.$$

If there exist constants m_p, M_p such that $m_p \leqslant \Delta^p a_n \leqslant M_p$, $n = 0, 1, 2, \ldots$, prove that

$$\frac{1}{p+1} m_p \leqslant \Delta^p A_n \leqslant \frac{1}{p+1} M_p, \qquad n = 0, 1, \ldots .$$

Solution by N. Sivakumar (student), University of Alberta, Edmonton, Canada. Let $k^{(p)}$ denote $k(k+1)\cdots(k+p-1)$. Since

$$(k+1)^{(p+1)} - k^{(p+1)} = (p+1)(k+1)^{(p)},$$

it follows that

$$\sum_{k=0}^{n} (k+1)^{(p)} = (n+1)^{(p+1)}/(p+1).$$

The desired bounds on the pth forward difference of the A's follow from the given uniform bounds on the pth forward differences of the a's and the following identity between these differences:

$$\Delta^p A_n = \frac{1}{(n+1)^{(p+1)}} \sum_{k=0}^{n} (k+1)^{(p)} \Delta^p a_k. \tag{$*$}$$

The identity $(*)$ may be proved by induction on p as follows.

For $p = 0$, $(*)$ holds by definition. If $(*)$ holds for a given value of p, then

$$\begin{aligned} \Delta^{p+1} A_n &= \Delta^p A_{n+1} - \Delta^p A_n \\ &= \frac{1}{(n+2)^{(p+1)}} \sum_{k=0}^{n+1} (k+1)^{(p)} \Delta^p a_k - \frac{1}{(n+1)^{(p+1)}} \sum_{k=0}^{n} (k+1)^{(p)} \Delta^p a_k \end{aligned}$$

so that

$$\begin{aligned}(n+1)^{(p+2)}\Delta^{p+1}A_n &= (n+1)\sum_{k=0}^{n+1}(k+1)^{(p)}\Delta^p a_k - (n+p+2)\sum_{k=0}^{n}(k+1)^{(p)}\Delta^p a_k\\ &= (n+1)^{(p+1)}\Delta^p a_{n+1} - (p+1)\sum_{k=0}^{n}(k+1)^{(p)}\Delta^p a_k\\ &= (n+1)^{(p+1)}\Delta^p a_{n+1} - \sum_{k=0}^{n}\left\{(k+1)^{(p+1)} - k^{(p+1)}\right\}\Delta^p a_k\\ &= \sum_{k=0}^{n}(k+1)^{(p+1)}\Delta^p a_{k+1} - \sum_{k=0}^{n}(k+1)^{(p+1)}\Delta^p a_k\\ &= \sum_{k=0}^{n}(k+1)^{(p+1)}\Delta^{p+1} a_k\end{aligned}$$

and $(*)$ holds with p replaced by $p+1$.

Also solved by A. Bege (Romania), I. E. Leonard (Canada), S. Marivani, R. K. Meany, J. E. Pečarić (Yugoslavia), University of South Alabama Problem Group, and the proposer.

Integral Quotients of Factorial Products

E 3107 [1985, 591]. *Proposed by Ira Gessel, Brandeis University.*

Show that if m and n are nonnegative integers, then $\dfrac{m!(2m+2n)!}{(2m)!n!(m+n)!}$ is an integer.

Solution I by A. A. Jagers, Universiteit Twente, The Netherlands.

$$\begin{aligned}\frac{m!(2m+2n)!}{(2m)!n!(m+n)!} &= \frac{2^n}{n!}(2m+2n-1)(2m+2n-3)\cdots(2m+1)\\ &= (-1)^n 2^{2n}\binom{-\frac{1}{2}-m}{n} = (-1)^n 2^{2n}\sum_{k=0}^{n}\binom{-\frac{1}{2}}{k}\binom{-m}{n-k}\\ &= \sum_{k=0}^{n} 2^{2n-2k}\binom{2k}{k}\binom{m+n-k-1}{n-k},\end{aligned}$$

a sum of integers.

Solution II by O. P. Lossers, Eindhoven University of Technology, The Netherlands. Denote the given quotient by $A(m,n)(m \geqslant 0,\, n \geqslant 0)$. Then we immediately see that

$$A(0,n) = \binom{2n}{n} \quad \text{is an integer for all } n \geqslant 0$$

$$A(m,0) = 1 \quad \text{is an integer for all } m \geqslant 0$$

Cube Inscribed in a Tetrahedron

E 3114 [1985, 666]. *Proposed by M. J. Pelling, London, England.* Find the largest cube that can be inscribed in some tetrahedron of volume 1.

Solution by J. G. Mauldon, Amherst College. A volume of $2/9$ can be achieved by considering the cube $0 \leqslant x, y, z \leqslant (2/9)^{1/3}$ in R^3 and the tetrahedron formed by the three coordinate planes and the plane $x + y + z = 6^{1/3}$; this is the greatest possible volume. More generally we have the following:

THEOREM. *If $V(T)$ is the volume of a tetrahedron T containing a given parallelopiped Π of volume $V(\Pi)$, then $V(T) \geqslant \frac{9}{2}V(\Pi)$.*

Similarly we have its two-dimensional analogue:

LEMMA. *If $A(\Delta)$ is the area of a triangle Δ containing a parallelogram R of area $A(R)$, then $A(\Delta) \geqslant 2A(R)$.*

The proof of the Theorem involves the Lemma, whose own independent proof is a simpler version of that for the Theorem.

Proof of the Lemma. By an affine transformation we may assume that R is a square of area a^2. Also we may assume that Δ lies in a compact disc, concentric with R, of radius $4a$, since otherwise Δ has a vertex D outside this disc, and the convex hull of $\{D\} \cup R$, a subset of Δ, already has area greater than $2a^2$. By compactness the continuous function $A(\cdot)$, defined on the set of such triangles Δ containing R, attains its infimum. Thus we may assume that $A(\Delta)$ is minimal.

If any one side f of Δ does not contain a side of R, then f meets R in only a single point P which must be a vertex of R and must also be the midpoint of f, since otherwise $A(\Delta)$ could be reduced by a small rotation around P of the side f alone.

Hence if no side of Δ contains a side of R then a side (in fact, each of two sides) of R is the join of the midpoints of two sides of Δ. Consequently, in this case, the opposite side of R lies in the third side of Δ, which contradicts the hypothesis.

The above contradiction shows that at least one side (the *base*) of Δ, of length b say, contains a side of R. If the corresponding altitude of Δ is of length $a + k$ ($k > 0$), then by similar triangles we have $k/a = (a + k)/b$ and so $A(\Delta) = \frac{1}{2}b(a + k) = a(a + k)^2/(2k) = 2a^2 + a(a - k)^2/(2k) \geqslant 2a^2 = 2A(R)$, as required.

Proof of the Theorem. This proof closely follows the proof of the Lemma. Similar argumentation shows that we may assume that Π is a cube C of volume a^3, that T is contained in a concentric compact ball of radius $27a$ and, by compactness and continuity, that $V(T)$ is minimal, so that in particular $V(T) \leqslant 9a^3/2$.

We denote by T^* the tetrahedron whose four vertices are the centroids of the four faces of T, observing that the faces of T^* are parallel to the faces of T and that T^* has one-third the linear dimensions of T, so that $V(T^*) = V(T)/27 \leqslant a^3/6$ by hypothesis and consequently, in particular, $V(T^*) \neq a^3/3$.

If any one face F of T does not contain an edge of C then F meets C in only a single point P which must be a vertex of C and must also be the centroid of F, since otherwise (cf. Pappus' Theorem) the (minimal) volume $V(T)$ could be reduced by some small rotation around P of the face F alone.

Suppose that no face of T contains an edge of C. Then all four vertices of the tetrahedron T^* are vertices of C, and, since $V(T^*) \neq a^3/3$, it follows that a face of T^* lies in a face of the cube C. (We have here used the fact that a tetrahedron whose vertices are four of the vertices of C has volume either $a^3/3$ or $a^3/6$, the latter case occurring only when a face of the tetrahedron lies in a face of C.) Hence the opposite face of the cube C lies in a face of T, which contradicts the hypothesis.

The above contradiction shows that at least one face F (the *base*) of T, of area $A(F)$, contains an edge of C and makes an angle θ $(0 \leqslant \theta \leqslant \pi/4)$ with a face of C. A plane parallel to F at a distance $a\cos\theta$ from F meets C in a rectangle R of area $a^2\sec\theta$ and (by the Lemma) meets T in a triangle Δ of area $A(\Delta) \geqslant 2a^2\sec\theta$.

If the altitude (to F) of T is of length $(k+a)\cos\theta$, where $k > 0$, we have (by similar tetrahedra) $A(F) = ((k+a)/k)^2 A(\Delta)$ and

$$
\begin{aligned}
V(T) &= \frac{1}{3}A(F)(k+a)\cos\theta = \frac{1}{3}A(\Delta)((k+a)/k)^2(k+a)\cos\theta \\
&\geqslant \frac{1}{3}(2a^2\sec\theta)(k+a)^3\cos\theta/k^2 \\
&= \frac{9a^3}{2} + \frac{a^2(4k+a)}{6k^2}(k-2a)^2 \\
&\geqslant \frac{9a^3}{2} = \frac{9}{2}V(\Pi)
\end{aligned}
$$

as required, with equality only if $k = 2a$ and $A(\Delta)$ is minimized in relation to the rectangle R contained in Δ.

Also solved by P. L. Hon (Hong Kong), L. Kuipers (Switzerland), J. Dou (Spain), and M. Orlowski & M. Pachter (South Africa).

Regular Simplexes in n Dimensions

E 3133 [1986, 132]. *Proposed by Peter D. Zvengrowski, The University of Calgary.*

Let $\Delta^n \subset \boldsymbol{R}^n$ be a regular n-simplex with center at 0. Let $\boldsymbol{a}_i = (a_{i1}, a_{i2}, \ldots, a_{in})$ be its vertices, $i = 0, 1, \ldots, n$. Show that for all j and k, $\Sigma_{i=0}^n a_{ij}a_{ik} = c\delta_{jk}$ for some constant c, where δ_{jk} is the Kronecker delta.

Solution by Bjorn Poonen, Winchester, MA. Without loss of generality, assume the length of each vector $\boldsymbol{a}_i$ is $\sqrt{n/(n+1)}$. By symmetry, $\boldsymbol{a}_i \cdot \boldsymbol{a}_j = \boldsymbol{a}_i \cdot \boldsymbol{a}_k$ for

$j,\ k \neq i$, and $\Sigma_{j=0}^{n}\boldsymbol{a}_j = \mathbf{0}$. Thus

$$0 = \boldsymbol{a}_i \cdot \sum_{j=0}^{n} \boldsymbol{a}_j = n/(n+1) + n(\boldsymbol{a}_i \cdot \boldsymbol{a}_k) \quad \text{and} \quad \boldsymbol{a}_i \cdot \boldsymbol{a}_k = -1/(n+1)$$

for all $i \neq k$. Now let $\boldsymbol{b}_i \in R^{n+1}$ be the vector formed by placing $\sqrt{1/(n+1)}$ after the n components of $\boldsymbol{a}_i$. Then $\boldsymbol{b}_i \cdot \boldsymbol{b}_k = \boldsymbol{a}_i \cdot \boldsymbol{a}_k + 1/(n+1) = \delta_{ik}$ for all $i \neq k$, so the $(n+1) \times (n+1)$ matrix whose ith column is $\boldsymbol{b}_i$ is orthogonal. But the transpose of an orthogonal matrix is orthogonal, so $\Sigma_{i=0}^{n} a_{ij} a_{ik} = \delta_{jk}$.

Similarly we can prove the following partial converse: If $\Sigma_{i=0}^{n} a_{ij} a_{ik} = c\delta_{jk}$ for some $c > 0$, and $\Sigma_{i=0}^{n}\boldsymbol{a}_i = \mathbf{0}$, then the endpoints of the vectors $\boldsymbol{a}_i$ form the vertices of a regular n-simplex with center at zero.

Proof. Without loss of generality assume $c = 1$. Define $\boldsymbol{b}_i$ as before. Then the $(n+1) \times (n+1)$ matrix whose ith column is $\boldsymbol{b}_i$ must be orthogonal, since the given conditions imply that its transpose is. Hence, $\delta_{ij} = \boldsymbol{b}_i \cdot \boldsymbol{b}_j = \boldsymbol{a}_i \cdot \boldsymbol{a}_j + 1/(n+1)$, so for all $i \neq j$ we have

$$\begin{aligned}(\boldsymbol{a}_i - \boldsymbol{a}_j)^2 &= \boldsymbol{a}_i \cdot \boldsymbol{a}_i - 2\boldsymbol{a}_i \cdot \boldsymbol{a}_j + \boldsymbol{a}_j \cdot \boldsymbol{a}_j \\ &= n/(n+1) - 2 \cdot 1/(n+1) + n/(n+1).\end{aligned}$$

But if the lengths of the sides of the simplex formed by the endpoints of the vectors $\boldsymbol{a}_i$ are equal then the simplex is regular, and its center is

$$1/(n+1) \sum_{i=0}^{n} \boldsymbol{a}_i = \mathbf{0}.$$

Also solved by O. P. Lossers (The Netherlands), D. B. Shapiro, A. Tissier (France), W. P. Wardlaw, J. B. Wilker, K. Zacharias (W. Germany), and the proposer.

E 3140 [1986, 216]. *Proposed by Khristo Boyadzhiev, University of Sofia, Bulgaria.* Let $K = \frac{1}{2}\ln^2(\sqrt{2} + 1)$.

(a) Show that

$$\int_1^\infty \frac{\ln\left(x + \sqrt{x^2 - 1}\right)}{x(1 + x^2)}\, dx = K.$$

(b) Also show that

$$\text{(i)} \qquad \int_1^\infty \frac{\arctan x}{x\sqrt{x^2 - 1}}\, dx = \frac{\pi^2}{8} + K$$

$$\text{(ii)} \qquad \int_0^1 \frac{\arctan x}{\sqrt{1 - x^2}}\, dx = \frac{\pi^2}{8} - K.$$

Editor's note. The statement of this problem gave a value of K equal to one half of what appears here. All the solvers noted the misprint.

Solution to part (a) by Hans Kappus, Rodersdorf, Switzerland. Denote the given integral by I. Substituting $x = \frac{1}{2}(t^{1/2} + t^{-1/2})$, we obtain

$$I = \int_0^1 \frac{t-1}{(t+1)(t^2+6t+1)} \ln t\, dt.$$

By partial fractions,

$$4I = \int_0^1 \left(\frac{2}{t+1} - \frac{1}{t+a} - \frac{1}{t+b} \right) \ln t\, dt,$$

where $a = 3 + 2\sqrt{2}$ and $b = 1/a = 3 - 2\sqrt{2}$. The substitution $u = bt$ gives

$$\int_0^1 \frac{\ln t}{t+a}\, dt = \ln a \ln(b+1) + \int_0^b \frac{\ln u}{u+1}\, du,$$

and an analogous equation holds with a and b interchanged. Hence,

$$4I = 2\int_0^1 \frac{\ln t}{t+1}\, dt - \ln a \ln(b+1) - \int_0^b \frac{\ln t}{t+1}\, dt - \ln b \ln(a+1) - \int_0^a \frac{\ln t}{t+1}\, dt$$

$$= \ln a \ln\left(\frac{a+1}{b+1}\right) - \int_1^a \frac{\ln t}{t+1}\, dt + \int_b^1 \frac{\ln t}{t+1}\, dt$$

$$= \ln^2 a - \int_1^a \frac{\ln t}{t+1}\, dt - \int_1^a \frac{\ln t}{t(t+1)}\, dt$$

$$\left(\text{substituting } \frac{1}{t} \text{ for } t \text{ in the second integral}\right)$$

$$= \ln^2 a - \int_1^a \frac{\ln t}{t}\, dt$$

$$= \frac{1}{2}\ln^2 a = 2\ln^2\sqrt{a} = 2\ln^2(\sqrt{2}+1).$$

Dividing by 4 completes the argument.

Solution to part (b) by the proposer. Consider the function

$$F(\alpha) = \int_1^{1/\sqrt{1-\alpha^2}} \frac{\arctan \alpha x}{x\sqrt{x^2-1}}\, dx, \qquad 0 \leqslant \alpha < 1.$$

By Leibniz's formula for differentiation,

$$F'(\alpha) = \frac{1}{\sqrt{1-\alpha^2}} \arctan \frac{\alpha}{\sqrt{1-\alpha^2}} + \int_1^{1/\sqrt{1-\alpha^2}} \frac{dx}{(1+\alpha^2x^2)\sqrt{x^2-1}}.$$

Put $t^2 = 1 - (1/x^2)$ and note that $\arcsin\alpha = \arctan\alpha/\sqrt{1-\alpha^2}$ to obtain

$$\begin{aligned}
F'(\alpha) &= \frac{\arcsin\alpha}{\sqrt{1-\alpha^2}} + \int_0^\alpha \frac{dt}{1+\alpha^2-t^2} \\
&= \frac{\arcsin\alpha}{\sqrt{1-\alpha^2}} + \frac{1}{2\sqrt{1+\alpha^2}}\ln\left(\frac{\sqrt{1+\alpha^2}+t}{\sqrt{1+\alpha^2}-t}\right)\Bigg|_0^\alpha \\
&= \frac{\arcsin\alpha}{\sqrt{1-\alpha^2}} + \frac{1}{2\sqrt{1+\alpha^2}}\ln\left(\frac{\sqrt{1+\alpha^2}+\alpha}{\sqrt{1+\alpha^2}-\alpha}\right) \\
&= \frac{\arcsin\alpha}{\sqrt{1-\alpha^2}} + \frac{\ln\left(\sqrt{1+\alpha^2}+\alpha\right)}{\sqrt{1+\alpha^2}}.
\end{aligned}$$

Integrating, we obtain

$$F(\alpha) = \frac{1}{2}(\arcsin\alpha)^2 + \frac{1}{2}\left(\ln\left(\alpha+\sqrt{1+\alpha^2}\right)\right)^2 + C.$$

Now $F(0) = 0$ implies that $C = 0$. Letting α approach 1 gives us the result in part (i).

We may derive the result in part (ii) from that in part (i) by noting that $\arctan x = \pi/2 - \arctan(1/x)$. Hence

$$\begin{aligned}
\int_0^1 \frac{\arctan x}{\sqrt{1-x^2}}\,dx &= \frac{\pi}{2}\int_0^1 \frac{dx}{\sqrt{1-x^2}} - \int_0^1 \frac{\arctan(1/x)}{\sqrt{1-x^2}}\,dx \\
&= \frac{\pi^2}{4} - \int_1^\infty \frac{\arctan t}{t\sqrt{t^2-1}}\,dt \\
&\quad (\text{setting } t = 1/x) \\
&= \frac{\pi^2}{8} - K.
\end{aligned}$$

Editorial comments. O. P. Lossers and the University of South Alabama Problem Group evaluated the integral in part (a) by residues. N. Ortner expressed the integrals as double integrals and evaluated them by interchanging the order of integration. The proposer solved part (a) by differentiating the function

$$F(\alpha) = \int_{1/\alpha}^\infty \frac{\ln\left(\alpha x + \sqrt{\alpha^2x^2-1}\right)}{x(1+x^2)}\,dx, \qquad \alpha > 0.$$

Most of the other solvers converted the integrals to infinite series. These series were summed most often by reference to handbooks, tables, or other sources.

Also solved by P. Bracken (Canada), K. W. Lau (Hong Kong), O. P. Lossers (The Netherlands), W. A. Newcomb, The University of South Alabama Problem Group, J. E. Wilkins, Jr., and J. W. Wrench. Part (a) was also solved by the Chico Problem Group, N. Ortner (Austria), K. Zackarias (West Germany) and the proposer. Part (b) was also solved by L. M. Christophe and A. J. Krishna (student).

ADVANCED PROBLEMS

6563. *Proposed by C. A. Spiro, SUNY at Buffalo.*

Prove that, if N is sufficiently large, at least one of the integers $N, N+1, \dots, N+9$ has more than two distinct prime factors.

6564. *Proposed by Gerald Weinstein, C.U.N.Y.*

Define $S(\theta) = \lim_{n \to +\infty} (\sin n)^{n^\theta}$.

(i) Show that $S(2)$ does not exist.

(ii) If C is a constant such that $|p/q - \pi| > q^{-C}$ for all pairs p, q of sufficiently large positive integers, prove that $S(\theta)$ exists and is zero for all $\theta > 2C - 2$.

Mahler [*Indagationes Math*, 15 (1953) 30–42] showed that we can take $C = 41.09$. For the best known value of C see D. V. Chudnovsky and G. V. Chudnovsky, *Springer Lecture Notes 1052*, (1984) 37–84.

6565. *Proposed by L. A. Rubel, University of Illinois at Urbana-Champaign.*

Suppose $A = \{a_1, a_2, a_3, \dots\}$ is a sequence of distinct positive integers and $B = \{b_1, b_2, b_3, \dots\}$ is another. Suppose A contains arbitrarily long arithmetic progressions.

(a) If $|b_n - a_n| \leq K$ for all n, where K is a fixed positive number, prove that B must contain arbitrarily long arithmetic progressions.

(b) If $\Sigma_{n=1}^{\infty} 1/\{1 + |b_n - a_n|\}$ diverges, does the same conclusion hold?

REVIEWS

EDITED BY JOSEPH KONHAUSER
Department of Mathematics, Macalester College, St. Paul, MN 55105

Enigmas of Chance: An Autobiography. By Mark Kac. Harper & Row, Publishers, New York, 1985, xxvii + 163 pp.

WOLFGANG FUCHS
Department of Mathematics, Cornell University, Ithaca, NY 14853

The Sloan Foundation had the splendid idea of asking eminent scientists to write their autobiographies, concentrating on their life in science, how it shaped their existence, what it means to them and what they see as its aim. *Enigmas of Chance* is the autobiography that Mark Kac wrote at the Foundation's request. It is a fitting memorial to a great mathematician and a fine person. Unfortunately, he did not live long enough to see its publication.

The manner of writing is delightful, witty and lucid, very much like Mark's* conversational style. Those who had the good luck of knowing him personally will hear his voice behind the printed page.

The book does many different things and can be enjoyed at many different levels.

It recalls the life of a brilliant man in a series of charming vignettes and delightful anecdotes. It is a (deservedly) happy story of success in his work and happiness with his family and friends. Mark's personality emerges with wonderful clarity, excepting only that his modesty did not let him dwell on some sides of it, like his extraordinary kindness and helpfulness to those in trouble, be they floundering graduate students or victims of totalitarian oppression.

Although Mark's life deserves to be called happy, it lay under the shadow of the Holocaust in which all his Polish family was wiped out. This makes his book a moving document of what it meant to be a Jewish scientist born in Poland, living through the dark times of our 20th century.

Mark tells vividly of his incurable infection by the mathematical virus while still a school boy, and gives an excellent account of the continuing affliction: The total absorption in a problem, the frustration when things do not work out, the elation when they do.

In keeping with the Sloan Foundation's charge, there is a very interesting description of Mark's mathematical evolution, a discussion about the nature of mathematics and even some actual mathematics (one chapter is "The Search for the Meaning of Independence"). Judging from a very small sample of nonmathematical

---

*I hope that the reader does not mind if I refer to my friend for the remainder of this review as Mark. (I do this despite my strong disapproval when hearing a well-known historian refer to Churchill as Winston during World War II. But that's different, isn't it?)

family and friends who have read the book, this chapter conveys the excitement and enthusiasm of discovery, even if the material itself remains inaccessible.

Since the general public (and expert opinion, too) is not even clear about what constitutes the subject matter of mathematics, the discussion of the nature of mathematics required great skill and careful selection. Mark steered clear of the philosophy of the subject and of foundational questions, except for a quick sneer at Bertrand Russell's well-known dictum that mathematics can be defined as the subject in which we never know what we are talking about, nor whether what we are saying is true. On the foundations Mark has this to say: "If a subject is robust, it should be insensitive to its foundations, and therefore too great an emphasis on the latter tends to produce a misleading slant. Concern for the foundations should come, if at all, after one has a firm intuitive grasp of the subject."

Mark stresses the fact that mathematics is a cultural activity as old as civilization itself. And he underlines the esthetic component of its appeal. He points out the two mainsprings of mathematical activity: The search for logical perfection and simplicity ("Pure Mathematics") and the quest for solutions of concrete problems, usually arising in other fields ("Applied Mathematics"). He disagrees sharply with those who only recognise one of these two parts as 'True Mathematics' and emphasizes the great strength that mathematics derives from its dual inspiration.

His own research is a perfect justification of this view. Even when working on abstract theories his starting point was always a concrete problem, simple enough to allow detailed calculations, complex enough to have the essential features of the general case. The ability to devise such problems is the hallmark of his genius. It was his firm conviction that all good work was founded on one simple, good, key idea, "the joker," as he called it. Once the joker was found, the rest followed, though it might still involve a lot of hard work. In this sense Mark was an 'applied mathematician,' no matter how abstract his subject matter might be. It is also true that in later life his interest shifted more and more towards physics. Yet it is my impression that his work in physics had an unmistakeable mathematical flavor, both in the choice of problems and in the standard of rigor.

Since it is not usual to write a book review with nothing but praise, let me finish by expressing regret at an omission in the book. In Mark's lifetime his special field, probability theory and its applications, grew explosively. To a large extent this is due to his own efforts. But there was also much work done by others. It would have been most interesting to have his comments on the progress achieved and on its history. He refrained from this completely; not even his famous colleague at Cornell, Will Feller, appears in the index. But, then, this is the life story of Mark Kac and it is quite proper that the wealth of his own ideas has pushed other topics aside.

I am sadly aware that my enumeration of the topics of *Enigmas of Chance* in no way recaptures the charm, the ebullience, and also the well-disguised seriousness of the book.

I urge the reader to rush out to get hold of a copy.

We all owe a great debt to the Sloan Foundation for godfathering this lovely work. What a pity that they were not around at the time of Georg Cantor or Dirichlet or Euclid.

Tilings and Patterns. By Branko Grünbaum and G. C. Shephard. W. H. Freeman & Co., New York, 1986. 700 + ix pages, 1406 illustrations, cloth, $59.95.

SOLOMON W. GOLOMB
Department of Mathematics, University of Southern California, Los Angeles, CA 90089

This is a marvelous book. I am tempted to say "I wish I had written it," but the comprehensiveness and attention to detail exceed anything I would have ever contemplated, let alone executed. The book is profusely illustrated, with an average of two figures per page, and each picture is meticulously drawn. It must have been a difficult decision for the publisher to confine all illustrations to black-and-white, thereby foregoing the art gift book market, though the production cost would no doubt have been exorbitant. However, for the mathematician, all the necessary distinctions are amply illustrated with shadings, letters, and numbers, and an excellent aesthetic effect is achieved.

One major achievement is the organization of the vast literature on planar tilings into a systematic structure, arranged into twelve chapters. The first eight chapters develop a "classical" theory of tilings, including classification of tilings with transitivity properties (Chapter 6), with respect to symmetries (Chapter 7), and when color is added (Chapter 8). Interesting recreational problems are liberally interspersed: rectangles of unequal squares, including Duijvestijn's decomposition of a square into 21 unequal squares (in section 2.4); superposition of two polygonal tilings to obtain a dissection of one polygon where the pieces can be reassembled to form the other polygon, in the fashion of H. Lindgren's 1972 Dover book on *Geometric Dissections* (in section 2.6); and many others.

The last four chapters contain some advanced material, mostly rather recent, and (to this reviewer, at least) especially interesting. Chapter 9, "Tilings by Polygons," includes a wealth of tiling material involving polyominoes and their cousins, polyiamonds and polyhexes, as well as a discussion of spiral tilings such as Voderberg's. Chapter 10, "Aperiodic Tilings," begins with "similarity tilings," including certain *rep-tiles*, which *can* tile aperiodically, and continues with the six-piece sets of Robinson, of Ammann, and of Penrose, which can *only* tile aperiodically. This topic is developed extensively, leading to the two-piece sets of Penrose (the kite-and-dart, and the Penrose rhombs) and Ammann's set (of two similar concave hexagons), which can tile the plane, but only aperiodically. The central unsolved problem is suitably highlighted: *Is there a single tile which can tile the plane aperiodically, but not periodically?* Chapter 11 ("Wang Tiles") develops Hao Wang's theory of the close relationship between Turing machines, undecidability, and the question of whether a given set of colored tiles (actually, "MacMahon squares") can tile the plane. The final chapter ("Tilings with Unusual Kinds of Tiles") discusses tiles with unconventional connectivity (e.g., with reentrant boundaries, and disconnected tiles), tilings by irregularly shaped figures, and tilings with special adjacency conditions on abutting tiles.

Sets of exercises are distributed liberally throughout the book, and it would be possible to use it as a text in a somewhat unconventional geometry course. The book

concludes with 42 pages of references, citing articles in sixteen languages, and a six-page index.

The authors have made a commendable effort to cite relevant references for each topic they discuss. My only complaint is that sometimes the *originator* of an idea is not directly credited, if the concept appears in the work of another author who in turn cites the inventor. Thus, Ammann is not listed as an author, which makes it difficult to attribute proper credit to his discoveries. Several of my own contributions (coining the term *rep-tile* and observing the aperiodic tilings they generate; the problem of tiling the plane using exactly one square of each integer side; etc.) are not directly attributed, as the reference given is to *Scientific American* columns by Martin Gardner, or to anthologies of such columns. (This may be mere petulance on my part. I *am* credited with inventing the term *polyominoes* in 1953, and with writing a principal reference on this subject in 1965.)

As extensive as this book is, it obviously cannot present every result ever obtained about tiling. However, the authors have excellent taste in selecting their material, and make an admirable effort to cite references to the topics which they do not cover. In a first edition of a book of this length there are inevitably some typographical errors, but they are surprisingly few and far between. I recommend this book enthusiastically to anyone interested in problems of tiling the plane.

TELEGRAPHIC REVIEWS

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T: Textbook	P: Professional Reading	1–4: Semesters
C: Computer Software	L: Undergraduate Library	**: Special Emphasis
S: Supplementary Reading	13: Grade Level	??: Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

General, S, P*. *Writing Mathematics Well: A Manual for Authors.* Leonard Gillman. MAA, 1987, ix + 49 pp, (P). [ISBN: 0-88385-443-0] A brief, pithy, blunt manual for writing mathematics; uses examples marked by checks ($\surd$) and bold crosses ($\times$) to distinguish right and wrong alternatives. Covers organization, mechanics (including copy-editing) as well as usage (mathematical English and standard English) from a traditional perspective (e.g., no mention of the tricky problems posed by non-sexist language or of the new editing, layout, and procedural choices posed by desk-top publishing). Too short to be comprehensive, but not too long to be read quickly and repeatedly. LAS

General, S, P, L. *No Way: The Nature of the Impossible.* Ed: Philip J. Davis, David Park. WH Freeman, 1987, xvi + 325 pp, $17.98. [ISBN: 0-7167-1813-8] A fascinating collection of essays by scientists, humanists, and artists on the importance and role of impossibility in various fields. An unusual approach to defining the limits of human achievement; provides a unique cross-disciplinary application of a fundamental concept in epistemology and mathematics. LAS

General, P, L. *Quiddities: An Intermittently Philosophical Dictionary.* W.V. Quine. Harvard U Pr, 1987, 249 pp, $20. [ISBN: 0-674-74351-2] Inspired by Voltaire's *Philosophical Dictionary,* the eminent philosopher Quine reflects on topics in and peripheral to philosophy, language and mathematics, including belief, communication, formalism, free will, Latin pronunciation, longitude and latitude, mathematosis (a disease), 'usage and absuage.' Whimsically humorous intellectual free play by a great thinker, for browsing. RB

Elementary, T(13: 1). *Algebra Programmed, Part 3, Third Edition.* Robert H. Alwin, Robert D. Hackworth, Joseph W. Howland. Prentice-Hall, 1987, xviii + 680 pp. [ISBN: 0-13-021940-1] Changes from the *Second Edition* (TR, March 1972) include less material on the structure of the real numbers and more emphasis on "skills and concepts involving exponents, radical expressions, and the graphing of linear and quadratic equations." JS

Elementary, T. *Intermediate Algebra, Fifth Edition.* Mervin L. Keedy, Marvin L. Bittinger. Addison-Wesley, 1987, xxvi + 646 pp, $28.95 (P) [ISBN: 0-201-15379-3]; *A Problem-Solving Approach to Intermediate Algebra, Second Edition,* 1986, x + 518 pp, $30.95 [ISBN: 0-201-12179-4]; *Intermediate Algebra with Problem Solving,* 1986, xii + 449 pp, (P). [ISBN: 0-201-15650-4] Three "new" entries in a plethora of new names and editions by the same authors. *Intermediate Algebra* features an "interactive worktext pedagogy;" changes in this "significant revision" of the *Fourth Edition* include a new focus on skill-maintenance; more applications, exercises, and examples. *A Problem-Solving Approach* covers linear equations earlier than usual as part of the emphasis on problem solving; no indication of changes from the *First Edition. Intermediate Algebra with Problem Solving* is a paperbound version of the previous text. Supplementary materials available for the series include instructional software, computerized testbank, videotape cassettes, and a placement test. JNC

Elementary, T(13: 1, 2). *Business Algebra & Mathematics.* Thomas F. George. Prentice-Hall, 1987, xiv + 634 pp. [ISBN: 0-13-091562-9] Algebraic topics up to quadratics and system of equations, the standard deviation and measures of central

tendancy, and eleven chapters on topics in business mathematics. Appendices on accounting and income taxes. FLW

Mathematics Appreciation, S*, L.** *Time Travel and Other Mathematical Bewilderments.* Martin Gardner. WH Freeman, 1987, ix + 295 pp, $12.95 (P); $17.95. [ISBN: 0-7167-1925-8; 0-7167-1924-X] Twelfth collection of reprints from Gardner's *Scientific American* column, this one devoted to paradoxes (nontransitivity, time travel), puzzles (card problems, rubber rope), pasttimes (tangrams, tilings, magic squares), and much more. Includes the famous 1975 April Fool's column that fooled thousands of readers with a counterexample to the four-color conjecture. LAS

Education, P. *No More Teacher Traps—Neither in the Name of Holmes Nor Carnegie.* Julius G. Goldberg. Vantage Pr, 1987, 410 pp, $18.95. [ISBN: 0-533-06550-X] Goldberg, a Russian-born and educated mathematics professor at Ohio State at Marion, addresses problems of declining class performance, mutually hostile student-teacher relations, grade inflation and other failings of the American educational system in six essays: aspects of educational reform; how to create a teacher, lesson, student; shortcomings of currently proposed reforms. RB

Education, S(16-17), P. *Stochastik in der Kollegstufe.* Harald Scheid. Lehrbrucher und Mono. zur Didaktik der Math., B. 6. Bibliographisches Institut, 1986, 250 pp, 32 DM (P). [ISBN: 3-411-03129-8] On probability and statistics in the secondary-school curriculum—what should be taught and how. JD-B

Education, L. *History in Mathematics Education.* Ed: I. Grattan-Guinness. Belin (8, rue Ferou, 75006 Paris, FRANCE), 1986, 208 pp, 60F (P). [ISBN: 2-905-746-05-X] Proceedings of a workshop held at the University of Toronto during the summer of 1983: a dozen papers by authors from around the world on historical themes (e.g., the origins of "proof") and on teaching practice (e.g., the use of transparencies). LAS

Education, S(15-16). *Teaching with Student Math Notes.* NCTM, 1987, iv + 123 pp, $12.50 (P). [ISBN: 0-87353-244-9] Collection of all the four-page Student Math Notes from the *NCTM News Bulletin* from September 1982 through May 1986. Additional teacher notes include teaching ideas, solutions, suggested extensions. Perforated and reproducible, suitable for teaching-training activities. MW

Education, S(11-13), L. *Enterprising Mathematics.* The Spode Group. Oxford U Pr, 1986, ix + 181 pp, $24.95 (P). [ISBN: 0-19-853652-6] An innovative collection of rich, realistic problems from practical life designed to motivate British high school age students who have been turned off by mathematics. Typical "problems" are projects with lengthy (page-long) descriptions requiring concrete mathematical exploration and analysis. Grouped by theme: sport, travel, parties, money, home, house design, buying, and "odds and ends." LAS

Education, S(15-18). *How To Evaluate Progress in Problem Solving.* Randall Charles, Frank Lester, Phares O'Daffer. NCTM, 1987, v + 85 pp, $5.60 (P). [ISBN: 0-87353-241-4] Practical advice for teachers based on authors' research. Discusses goals of teaching problem solving and describes evaluation techniques of observation and questioning, student self-assessment, holistic scoring as well as the familiar multiple choice and completion tests. Tips on organizing evaluation and using results. Bibliography of problem-solving resources. Excellent for in-service programs. MW

Education, S(15-16). *How to Teach Mathematics Using a Calculator: Activities for Elementary and Middle School.* Terrence G. Coburn. NCTM, 1987, v + 58 pp, $6.90 (P). [ISBN: 0-87353-245-7] Reproducible activity sheets covering topics from numeration to prealgebra. A few straightforward "check your answer on the calculator" types; most involve pattern finding and problem solving. Useful in elementary education. MW

Education, S(13-16). *Problem-solving Techniques Helpful in Mathematics and Science.* Charles A. Reeves. NCTM, 1987, iii + 35 pp, $4 (P). [ISBN: 0-87353-246-5] Designed to familiarize elementary teachers with common problem-solving strategies, such as working backwards, making a chart or diagram, search for patterns on guess-check-revise. Emphasis on mathematics problems. Useful as overview, but neophytes would need more problems. MW

History, S, P, L*. *Women of Mathematics: A Biobibliographic Sourcebook.* Ed: Louise G. Grinstein, Paul J. Campbell. Greenwood Pr, 1987, xx + 292 pp, $45. [ISBN: 0-313-24849-4] 43 profiles of women (either deceased or born before 1925) who overcame familial and societal pressures to achieve distinction as mathematicians. Each profile contains a biography, a summary of professional work (in nontechnical terms), a bibliography of writings by and about the subject. Profiles are written by a wide variety of mathematicians, historians, and scholars. A distinctive and valuable resource for anyone interested in equality of opportunity in mathematics. LAS

History, P, L. *Riemann, Topology, and Physics.* Michael Monastyrsky. Transl: James King, *et al.* Birkhauser Boston, 1987, xiii + 158 pp, $39.50. [ISBN: 0-8176-3262-X] Translations of two works by a Russian physicist: a biography of Riemann, and a monograph on topological methods in physics. The biography "describes in few words and fewer equations the revolutionary ideas which Riemann brought in to mathematics and physics ...." In the monograph, homotopy theory provides a unified descrip-

tion of physical phenomena, including defects in liquid crystal, monopoles in gauge fields. RB

History, P, L. *Kepler's Physical Astronomy*. Stud. in Hist. of Math. & Phys. Sci., V. 13. Springer-Verlag, 1987, vii + 216 pp, $58. [ISBN: 0-387-96541-6] Critical examination of Kepler's *Astronomia nova*, a record of his ten-year labors on pre-Newtonian astronomy written to convince trained astronomers that the then-current planetary theory was wrong. Dispels notion that Kepler was merely "a diligent condenser of data, or a numerically inclined mystic." JK

History, L. *The Mechanization of the World Picture: Pythagoras to Newton*. E.J. Dijksterhuis. Transl: C. Dikshoorn. Princeton U Pr, 1986, x + 537 pp, $14.50 (P). [ISBN: 0-691-02396-4] Reprinting of the 1961 Oxford University Press edition (a translation of the Dutch work *De Mechanisering van het Wereldbeeld*) with a new foreword by D.J. Struik. Detailed, scholarly account of the development of a mechanistic world view. Written for the general reader, "who is assumed to have no special training in mathematics or physics." BH

History, P, L*. *The Newton Handbook*. Derek Gjertsen. Routledge & Kegan Paul (US Distr: Methuen), 1986, xiv + 665 pp, $59.95. [ISBN: 0-7102-0279-2] An alphabetical mini-encyclopedia of Newton: his theories, friends, relatives, publications, intellectual descendents, and cultural influences, all arranged alphabetically (whence ... Bible, Bibliographies, Binomial Theorem, ...). Includes much of interest to anyone wanting to comprehend the incredible depth and diversity of Newton's accomplishments. Contains thorough summaries and complete bibliographic details on Newton's voluminous published and unpublished manuscripts. LAS

History, T*(16), S. *The Origins of the Infinitesimal Calculus*. Margaret E. Baron. Dover, 1987, viii + 304 pp, $7.95 (P). [ISBN: 0-486-65371-4] Thorough study of the rise of geometric techniques from the Greeks to early 17th century. Excellent detail and documentation allowing reader to work through many problems and to find original sources. Would serve well as a text for undergraduate course in history of calculus. Dover reprint of 1969 Pergamon Press edition (TR, October 1969). MR

History, P, L. *The Historical Development of Quantum Theory, Volume 5: Erwin Schrödinger and the Rise of Wave Mechanics, Part 1: Schrödinger in Vienna and Zurich, 1887-1925*. Springer-Verlag, 1987, xviii + 366 pp, $54. [ISBN: 0-387-96284-0] Covers in two parts Schrödinger's path to (but not including) his fundamental papers of 1926 on wave mechanics. First part is devoted to his early life and work in Vienna. Second part concentrates on Schrödinger's post-World War I work, mainly in statistical mechanics, and the work of others on the development of quantum mechanics and theory of atomic structure. MR

Logic, P. *Six Papers in Logic*. S.N. Artemov, *et al.* AMS Transl. Ser. 2, V. 135. AMS, 1987, ix + 79 pp, $38. [ISBN: 0-8218-3111-9] Topics are: finite model for finitely axiomatizable theory that is categorical in all infinite powers, spectrum function of countable theory, Gentzen systems of postulates for set theory, modal logic and extensions of arithmetic, complexity of disjunctive normal forms. KS

Logic, T(18: 1), P. *Multiple Forcing*. T. Jech. Tracts in Math., V. 88. Cambridge U Pr, 1986, viii + 136 pp, $34.50. [ISBN: 0-521-26659-9] Set theorists use forcing to construct models of set theory. This book presents unified treatment of product, iterated and proper forcing. Includes applications to algebra and topology. Assumes first graduate course in set theory. KS

Foundations, P, L*. *Reflections on Kurt Gödel*. Hao Wang. MIT Pr, 1987, xxvi + 336 pp, $25. [ISBN: 0-262-23127-1] A detailed commentary on "a life of fundamental theoretical work," covering Gödel's wide ranging contributions to mathematics, philosophy of science, and metaphysics, as well as details of events and interests that marked his life. Written in diary style, mixing abbreviated notes (e.g., "G," rarely "Gödel") with reflective commentary on the biographer's own views and relations with Gödel. A fascinating contribution to intellectual history, revealing unknown depth and richness to Gödel's thought. LAS

Combinatorics, P. *A Source Book in Matroid Theory*. Joseph P.S. Kung. Birkhauser Boston, 1986, 413 pp, $59.50. [ISBN: 0-8176-3173-9] An anthology of papers, both classical and recent, on matroid theory. The papers are preceded by commentary providing background, notation, and historical notes. Suitable as an introduction to matroid theory. LC

Discrete Mathematics, T(13-14: 1), L. *A Primer of Discrete Mathematics*. Daniel T. Finkbeiner II, Wendell D. Lindstrom. Books in the Math. Sci. WH Freeman, 1986, xvi + 363 pp. [ISBN: 0-7167-1815-4] True to its name, the book is a carefully-written introduction to several basic concepts presented in interesting and challenging ways. Includes elementary set theory, proofs, functions, relations, combinatorial methods (including generating functions), and graph theory with applications. Exercises, answers. JS

Number Theory, P. *Elliptic Functions and Rings of Integers*. Ph. Cassou-Noguès, M.J. Taylor. Prog. in Math., V. 66. Birkhauser Boston, 1987, xvii + 198 pp, $29.50. [ISBN: 0-8176-3350-2] Covers recent work on the structure of the rings of integers of Abelian extensions of quadratic imaginary number fields. The approach is based on complex function

theory and elliptic functions, instead of algebraic geometry and elliptic curves. RH

Number Theory, T*(16: 1), S, L. *Algebraic Number Theory, Second Edition.* Ian Stewart, David Tall. Math. Ser. Chapman & Hall, 1987, xix + 262 pp, $19.95 (P). [ISBN: 0-412-29690-X] Changes since *First Edition* (TR, May 1980) include corrections, historical updates on Fermat's Last Theorem and the class number problem, minor deletions, and two new, short sections on computing rings of integers and on factorization in cyclotomic fields (neither referred to in Index). Remains a good book. GG

Number Theory, P. *Lecture Notes in Mathematics-1252: Représentations de Weil et GL_2 Algèbres de division et GL_n.* Tetsuo Kaise. Springer-Verlag, 1987, vii + 203 pp, $20 (P). [ISBN: 0-387-17827-9] A report on the author's effort, carried out over several years, to develop non-abelian class field theory. SG

Linear Algebra, T(17: 1), S. *Matrix Theory: A Second Course.* James M. Ortega. Univ. Ser. in Math. Plenum Pr, 1987, xii + 262 pp, $34.50. [ISBN: 0-306-42433-9] Developed from lecture notes for in-coming applied mathematics graduate students; review of matrix theory, linear spaces and operators, and canonical forms followed by chapters on quadratic forms and optimization, differential and difference equations, and miscellaneous topics. JNC

Group Theory, P. *Combinatorial Group Theory and Topology.* Ed: S.M. Gersten, John R. Stallings. Annals of Math. Stud., No. 111. Princeton U Pr, 1987, viii + 551 pp, $65; $22.50 (P). [ISBN: 0-691-08409-2; 0-691-08410-6] Proceedings of a Alta, Utah conference (July 1984) which encompasses aspects of group theory arising from low-dimensional topology. Twenty-five papers, including numerous open problems and reports on current research. RB

Group Theory, P. *Subnormal Subgroups of Groups.* John C. Lennox, Stewart E. Stonehewer. Math. Mono. Clarendon Pr, 1987, xiii + 253 pp, $69. [ISBN: 0-19-853552-X] An up-to-date survey of the subject beginning with Wielandt's 1939 paper. Contains many results on the joining of subnormal subgroups, criteria for subnormality, and characterizations of groups satisfying various subnormality conditions. SG

Group Theory, T(16), S. *Introduction to Group Characters, Second Edition.* Walter Ledermann. Cambridge U Pr, 1987, x + 227 pp, $49.50; $14.95 (P). [ISBN: 0-521-33246-X; 0-521-33781-X] Changes from the *First Edition* (TR, June-July 1978) include a treatment of arithmetic properties of characters and the Frobenius-Schur condition for real representations. Other notable topics include induced characters, permutation groups, and the Burnside pq-theorem. SG

Group Theory, P. *Free Group Rings.* Ed: Narain Gupta. Contemp. Math., V. 66. AMS, 1987, xi + 129 pp, $22 (P). [ISBN: 0-8218-5072-5] Free group rings are the group rings over Z of free groups. This research monograph systematically presents some developments in the subject since Magnus' work (1930's); intended for workers in combinatorial group theory. Introduction; the Magnus embedding; the Fox problem and its recent solution; the Dimension Subgroup problem; topics and open problems. RB

Group Theory, S(18), P. *Proceedings of Groups—St. Andrews 1985.* Ed: E.F. Robertson, C.M. Campbell. London Math. Soc. Lect. Note Ser., V. 121. Cambridge U Pr, 1986, ix + 358 pp, $39.50 (P). [ISBN: 0-521-33854-9] A selection of 37 papers chosen from the presentations given at the international conference on groups held at St. Andrews, Scotland in the summer of 1985. Includes survey articles by Bachmuth, Baumslag, Neumann, Roseblade, and Tits. JS

Algebra, P. *Semigroups and Their Applications.* Ed: Simon M. Goberstein, Peter M. Higgins. D Reidel (US Distr: Kluwer Academic), 1987, xiv + 211 pp, $69. [ISBN: 90-277-2463-6] A collection of over 20 papers on a variety of topics in semigroup theory presented at a conference held at Chico State in 1986. SG

Algebra, P. *Antimorphic Action: Categories of Algebraic Structures with Involutions or Anti-Endomorphisms.* W.H. Cornish. Res. & Expos. in Math., V. 12. Heldermann Verlag, 1986, xix + 189 pp, $30 (P). [ISBN: 3-88538-212-1] A categorical approach to certain involutions of a Lie algebra, a quadratic Jordan algebra, or a free lattice in a self-dual variety of lattices. Explores properties of the variety formed by those algebras in a variety of finitary algebras expanded by antimorphisms which are acted upon by members of a fixed monoid in a homomorphic or antihomomorphic manner. The deepest results are relevant to algebras obtained from propositional logics. LW

Algebra, S(18), P. *Dimensions of Ring Theory.* Constantin Năstăsescu, Freddy van Oystaeyen. Math. & Its Applic. Kluwer Academic, 1987, xi + 360 pp, $74. [ISBN: 90-277-2461-X] Treatment of several of the concepts of dimension in ring theory, including Krull and Gabriel dimension, homological dimension, and Gelfand-Kirillov dimension. Exercises, bibliographical comments by chapter; references; index. JS

Algebra, S(18), P. *The Algebraic Structure of Crossed Products.* Gregory Karpilovsky. Math. Stud., V. 142. Elsevier Science, 1987, x + 348 pp, $71 (P). [ISBN: 0-444-70239-3] Assumes basic algebra background leading to a discussion of various questions for cross products. Chapters on group-graded algebras, classical theory (the Brauer group), Clifford theory for graded algebras, primitive and prime ideals. Bibliography, index. JS

Algebra, P. *Near-Rings and Near-Fields.* Ed: Gerhard Betsch. Math. Stud., V. 137. Elsevier Science, 1987, xiv + 298 pp, $62.25 (P). [ISBN: 0-444-70191-5] Proceedings of an August 1985 conference held at the Mathematical Institute of the University of Tübingen. Consists of two invited lectures and thirty additional papers. Focuses on general structure of near-rings, which may have a non-commutative addition and only a one-sided distributive law, but satisfy the other ring axioms, and of near-fields, near-rings whose non-zero elements form a group under multiplication. GG

Calculus, T(14: 1). *Vector Calculus.* Peter Baxandall, Hans Liebeck. Appl. Math. & Comput. Sci. Ser. Clarendon Pr, 1986, ix + 550 pp, $49.95 (P). [ISBN: 0-19-859652-9] A first-year text covering the usual topics—functions of several variables, differentiation, integration, including the theorems of Green, Stokes, and Gauss. A comfortable familiarity with basic analysis and linear algebra is assumed, although reviewed in the first chapter. A straightforward, concise introduction to vector calculus done with more rigor than usual. Many good examples illustrating important topics. Nice use of counterexamples in exercises to show significance of theorems' hypotheses. PS

Calculus, T(13-14: 2). *Höhere Mathematik I: Eine Einführung.* Günter Hellwig. Bibliographisches Institut, 1986, 411 pp, 29,80 DM (P). [ISBN: 3-411-05553-7] *Second Edition;* first published in 1971. Covers largely traditional topics in calculus of functions of one variable. However, it begins with fairly detailed treatment of real numbers, defines sine and cosine as solutions of differential equations, and ends with sketch of development of Riemann-Stieltjes integral. JD-B

Calculus, S*(13), L. *Motivating A-Level Mathematics: A Source Book.* The Spode Group. Oxford U Pr, 1986, ix + 130 pp, $24.95 (P). [ISBN: 0-19-853653-4] A series of case studies in applications of A-level mathematics (functions, calculus, mechanics, probability, and statistics) intended to motivate British vocational students between 16 and 19 years old. Well-known examples (radioactive decay) mix with unconventional applications (screen slopes, tennis service, blood donors). Each case study concludes with a few exercises, whose solutions occupy the final quarter of the book. LAS

Complex Analysis, T*(17-18: 1, 2), S*, P. *Lectures on Complex Approximation.* Dieter Gaier. Birkhauser Boston, 1987, xv + 196 pp, $42. [ISBN: 0-8176-3147-X] Translation by Renate McLaughlin of the 1980 *Vorlesungen über Approximation in Komplexen,* with some updates. Part I is constructive, a classical treatment of series expansions and interpolation. Part II covers the more modern existence theory. Both parts are largely self-contained, yet give statements (with references) of more recent developments. Extensive bibliography. GG

Complex Analysis, P. *Lecture Notes in Mathematics-1268: Complex Analysis.* Ed: S.G. Krantz. Springer-Verlag, 1987, 195 pp, $20 (P). [ISBN: 0-387-18094-X] Proceedings of a conference on complex analysis held at Penn State in March, 1987. The lead article is an account of some of the progress in several complex variables over the past fifteen years. RH

Differential Equations, T(15-16: 1), S, L. *Nonlinear Ordinary Differential Equations, Second Edition.* D.W. Jordan, P. Smith. Appl. Math. & Comput. Sci. Ser. Clarendon Pr, 1987, ix + 381 pp, $24.95 (P); $42.50. [ISBN: 0-19-859656-1; 0-19-859657-X] Introduction to dynamical systems on intuitive level. Emphasis on qualitative approach. Flexible. Many chapters self-contained. Emphasis on applications with examples in mechanics, electrical systems, and biology. Changes include many chapters revised or rewritten, added up-to-date material on nonlinear systems and new elementary treatment of bifurcation, structural stability, and chaotic behavior of oscillators. Well-written. Good exercises, many with answers. (*First Edition,* TR, June-July 1978.) JK

Differential Equations, P. *Lecture Notes in Mathematics-1258: Spectral Theory of Ordinary Differential Operators.* Joachim Weidmann. Springer-Verlag 1987, vi + 303 pp, $28.60 (P). [ISBN: 0-387-17902-X] General theory of ordinary differential equations covering operators of arbitrary order n operating on C^m-valued functions for arbitrary m. Applications to Sturm-Liouville operators and Dirac systems. Some new results "connected with existence of self-adjoint realizations with separate boundary conditions, multiplicity of the spectrum and the absolute continuity of the spectrum." Fairly self-contained. For mathematicians interested in applications to physics and to physicists with mathematical background. JK

Differential Equations, P*. *Linear Differential Equations and Group Theory from Riemann to Poincaré.* Jeremy Gray. Birkhauser Boston, 1986, xxv + 460 pp, $51.50. [ISBN: 0-8176-3318-9] Historical origins of the geometric view of complex function theory culminating in Klein's and Poincaré's theory of automorphic functions. Tracing of two converging approaches both with roots in the theory of elliptic functions, one via the hypergeometric differential equation and the other via Galois theory of polynomial equations. Extensive notes and references, often to original sources. JK

Differential Equations, T(14-15: 2). *Analysis III: Gewöhnliche Differentialgleichungen und Ausbau der Infinitesimalrechnung.* Erich Martensen. Bibliographisches Institut, 1986, x + 242 pp, 26,80 DM (P). [ISBN: 3-411-06834-5]

Differential Equations, P. *Asymptotic Methods for Relaxation Oscillations and Applications.* Johan Grasman. Appl. Math. Sci., V. 63. Springer-Verlag, 1987, xiii + 221 pp, $28 (P). [ISBN: 0-387-96513-0] Relaxation oscillation refers to a nonlinear type of oscillation often encountered in physics and biology in which certain systems behave as if periodically reset in an initial state. Grasman thoroughly covers the method of matched asymptotic expansions to approximate orbits of relaxation oscillators. Begins with Van der Pol's equation and gives special attention to biological and chemical relaxation oscillators. Sections on free oscillation, forced oscillation and mutual entrainment, and the Van der Pol oscillator with a sinusoidal forcing term. With exercises. PS

Differential Equations, T(14-15: 1). *A First Course in Differential Equations with Applications, Third Edition.* William R. Derrick, Stanley I. Grossman. West, 1987, xv + 545 pp, $29.76. [ISBN: 0-314-26895-2] First-order equations, second- and higher-order linear equations, power series solutions, Laplace transforms, linear systems, numerical methods, introduction to Fourier series and partial differential equations. Many sections on applications; "excursions" into non-traditional topics. Almost three hundred examples and 2000 exercises, many of an applied nature. DFA

Differential Equations, P, L. *Inverse Spectral Theory.* Jürgen Pöschel, Eugene Trubowitz. Pure & Appl. Math., V. 130. Academic Pr, 1987, x + 192 pp, $29.95. [ISBN: 0-12-563040-9] Can you hear the density of a string? In other words, how many ways can mass be distributed along a vibrating string to produce a given set of frequencies? This is a physical example that leads to a question in inverse spectral theory. The book is based on lectures given at the Courant Institute in 1980 and is intended for a general mathematical audience. It discusses the problem of characterizing the Sturm-Liouville equations that produce a given set of eigenvalues under Dirichlet boundary conditions. AM

Differential Equations, T(14-15: 1). *Introduction to Differential Equations with Boundary Value Problems, Third Edition.* William R. Derrick, Stanley I. Grossman. West, 1987, xv + 630 pp, $32.16. [ISBN: 0-314-26897-9] The authors' *A First Course in Differential Equations with Applications, Third Edition,* with additional material on Fourier series and partial differential equations, a chapter on nonlinear equations and stability (concerning two-dimensional systems), and examples and exercises to go with these extra topics. DFA

Partial Differential Equations, T(18), L. *Partial Differential Equations: An Introduction to a General Theory of Linear Boundary Value Problems.* Aleksei A. Dezin. Transl: Ralph P. Boas. Springer-Verlag, 1987, 163 pp, $82. [ISBN: 0-387-16699-8] The point of view in this text is how the boundary conditions influence the solvability of linear partial differential equations; the basic approach is through linear operators in Hilbert space. LC

Partial Differential Equations, P. *The Equations of Navier-Stokes and Abstract Parabolic Equations.* Wolf von Wahl. Aspects of Math., V. E8. Friedr Vieweg & Sohn, 1985, xxiv + 264 pp, $25 (P). [ISBN: 3-528-08915-6] Functional analytic consideration of instationary Navier-Stokes equations. Studies admissible initial data and global continuation of local solutions. Applications to global solvability of semilinear parabolic systems. DFA

Partial Differential Equations, P. *Lecture Notes in Mathematics-1270: Nonlinear Hyperbolic Problems.* Ed: C. Carasso, P.-A. Raviart, D. Serre. Springer-Verlag, 1987, xv + 351 pp, $32.90 (P). [ISBN: 0-387-18200-4] Twenty-five papers from an international conference held in Saint-Etienne on January 13-17, 1986, addressing open questions, existence and qualitative theory for hyperbolic partial differential equations; and general theory and main applications in numerical analysis of the same. DFA

Numerical Analysis, T(15-16: 1), L. *Applied Numerical Methods for Digital Computation, Third Edition.* M.L. James, G.M. Smith, J.C. Wolford. Harper & Row, 1985, xiii + 753 pp. [ISBN: 0-06-043281-0] Roots of equations, systems of equations, curve fitting, integration and differentiation, initial- and boundary-value problems for ordinary differential equations, introduction to partial differential equations. First chapter presents Fortran, last chapter introduces IBM's simulation language CSMP. Flowcharts and programs throughout. Exercises are for solution using computer programs or (in some cases) pocket calculators. DFA

Numerical Analysis, T(15-16: 1, 2). *Introduction to Numerical Analysis, Second Edition.* F.B. Hildebrand. Dover, 1987, xiii + 669 pp, $13.95 (P). [ISBN: 0-486-65363-3] Unabridged, slightly corrected republication of the 1974 McGraw-Hill edition (TR, November 1974). DFA

Numerical Analysis, P. *Tables for Converting Polynomials and Power Series into Chebyshev Series.* Herbert E. Salzer, Norman Levine. Applied Science, 1984, 65 pp, $12 (P). [ISBN: 0-915061-01-5] Tables of coefficients, computed to 35 significant figures, for converting polynomials in x of degree n, which includes truncated power series, into Chebyshev series, through $n = 50$ for x in the interval [0,1] and through $n = 100$ for x in the interval [-1,1]. Includes a brief discussion of Chebyshev series, information about incorporating the tables in computer programs, and additional references. RH

Numerical Analysis, T(17-18: 1). *Nonlinear Methods in Numerical Analysis.* Annie Cuyt, Luc Wuytack. Math. Stud., V. 136. Elsevier Science,

1987, 278 pp, $53.25 (P). [ISBN: 0-444-70189-3] An introduction to several nonlinear techniques for the solution of various numerical problems. The techniques result mainly from the use of Padé approximants and rational interpolants. The text is fairly self-contained. Each chapter includes a section of remarks suggesting extensions of the theory and a problem section. RH

Numerical Analysis, T(16-18: 2). *Numerical Algorithms.* Ed: J.L. Mohamed, J.E. Walsh. Clarendon Pr, 1986, xii + 356 pp, $60. [ISBN: 0-19-853364-0] Presents a wide spectrum of computational problems, for example eigenvalue problems, solution of differential equations, and curve fitting, and the algorithms used to solve them. Includes both the theoretical and practical ideas needed to design and implement the algorithms. Gives descriptions and sources of the current numerical software for the methods described. A good reference text for both specialists and non-specialists, or possibly a text for advanced undergraduate or graduate courses in computational mathematics. RH

Numerical Analysis, P. *Numerical Integration: Recent Developments, Software and Applications.* Ed: Patrick Keast, Graeme Fairweather. NATO ASI Ser C, V. 203. D Reidel (US Distr: Kluwer Academic), 1987, xiii + 394 pp, $89. [ISBN: 90-277-2514-4] A collection of twenty-five papers and ten extended abstracts of papers presented at the NATO Advanced Research Workshop which was held at Dalhousie University, Halifax, Canada, August 11-15, 1986. RH

Numerical Analysis, S(16-17), P. *Numerische Mathematik für Ingenieure.* Gisela Engeln-Müllges, Fritz Reutter. Bibliographisches Institut, 1987, xvi + 524 pp, 48 DM (P). [ISBN: 3-411-03151-4]

Functional Analysis, P. *Lecture Notes in Mathematics-1242: Functional Analysis II.* Ed: S. Kurepa, H. Kraljević, D. Butković. Springer-Verlag, 1987, vii + 432 pp, $37.10 (P). [ISBN: 0-387-17833-3] Lecture notes from a conference at the Inter-University Center of Postgraduate Studies, Dubrovnik, Yugoslavia, November 1985. Centers mainly on functional-analytic aspects of probability and operator theory on Hilbert and Banach spaces. LC

Functional Analysis, P. *Saks Spaces and Applications to Functional Analysis, Second, Revised Edition.* J.B. Cooper. Math. Stud., V. 139. Elsevier Science, 1987, ix + 372 pp, $75.50 (P). [ISBN: 0-444-70219-9] Closely related to the theory of locally convex spaces and generalized Banach spaces, mixed spaces—or Saks spaces—are vector spaces with a norm and a locally convex topology. Theory of Saks spaces is related to special topologies used on spaces of functions where natural norm topologies are not suitable for certain applications. This *Second Edition* takes recent developments in functional analysis into account. (1978 North-Holland edition, TR, November 1978.) AWR

Functional Analysis, P. *Characterizations of Inner Product Spaces.* Dan Amir. Oper. Theory, V. 20. Birkhauser Boston, 1986, 200 pp, $31.50. [ISBN: 0-8176-1774-4] A catalog of characterizations of inner-product spaces by normed-space geometry and approximation theoretic properties. Arguments presented are mostly elementary and often geometric. A reference for the expert which is also accessible to advanced undergraduates. LW

Functional Analysis, P. *On Banach Algebras, Renewal Measures and Regenerative Processes.* J.B.G. Frenk. CWI Tract, V.38. Math Centrum, 1987, 201 pp, Dfl. 31.60 (P). [ISBN: 90-6196-321-4] Formally, tersely, yet clearly written. Presents a characterization of the space of homomorphisms on the Banach algebra of complex measures concentrated on the positive half-line. Results on Banach algebras are then used to derive asymptotic results for the renewal sequence and the renewal measure in the case that the expectation of the waiting time distribution is finite. LW

Analysis, T(15-16: 1, 2). *Elements of Real Analysis.* David A. Sprecher. Dover, 1987, viii + 343 pp, $8.95 (P). [ISBN: 0-486-65385-4] A republication of the 1970 Academic Press edition (TR, April 1972). LC

Analysis, P. *Transactions of the Fourth Army Conference on Applied Mathematics and Computing.* US Army Research Office (PO Box 12211, Research Triangle Park, NC 27709), 1986, xxx + 1345 pp, (P). One hundred technical papers from a conference held May 27-30, 1987 at Cornell University as part of the opening of Cornell's new Mathematical Sciences Institute (MSI). Includes papers on combustion, fluid dynamics, stochastic analysis, parallel computation, numerical analysis. LAS

Algebraic Geometry, P. *Current Trends in Arithmetical Algebraic Geometry.* Ed: Kenneth A. Ribet. Contemp. Math., V. 67. AMS, 1987, x + 293 pp, $30 (P). [ISBN: 0-8218-5070-1] Six selected talks plus four additional papers from the 1985 Joint Summer Research Conference in Algebraic Geometry held in Arcata, California. Main topics involve cohomology, hyperbolic geometry, and Arakelov theory. GG

Algebraic Geometry, T(17: 4), S, P.** *Plane Algebraic Curves.* Egbert Brieskorn, Horst Knörrer. Transl: John Stillwell. Birkhauser Boston, 1986, vi + 721 pp, $49. [ISBN: 0-8176-3326-X] Wonderful introduction to the investigation of singularities of algebraic curves. Topics covered range from elementary methods (e.g., decomposition into irreducible components and Bézout's theorem) to resolution of singularities (e.g., localisation and topology of singularities). Along the way the author includes hundreds of drawings and also historical, heuristic, and method-

ological considerations, topics often neglected in the modern rush towards proving more theorems. MR

Algebraic Geometry, P. *Lecture Notes in Mathematics-1239: Diophantine Approximations and Value Distribution Theory.* Paul Vojta. Springer-Verlag, 1987, x + 132 pp, $12.80 (P). [ISBN: 0-387-17551-2] A summary of the author's research on the question of when algebraic varieties of dimension two or more have a finite number of rational points over every number field. SG

Algebraic Geometry, P. *Lecture Notes in Mathematics-1269: Nash Manifolds.* Masahiro Shiota. Springer-Verlag, 1987, vi + 223 pp, $20 (P). [ISBN: 0-387-18102-4] A study of C^r Nash manifolds which are finite semialgebraic open sets glued together by certain diffeomorphisms. SG

Differential Geometry, P. *Lectures on Fibre Bundles and Differential Geometry.* J.L Koszul. Springer-Verlag, 1986, ii + 127 pp, $12.60 (P). [ISBN: 0-387-12876-X] These are the well-known notes of Koszul's 1960 lectures at the Tata Institute for Fundamental Research, Bombay, re-issued (the original typewriting, but improved paper). Algebraic (non-manifold) formulation of covariant derivation and related notions; connection forms on principal bundles; relations between covariant derivations and connection forms; connections in almost complex and holomorphic bundles. RB

Differential Geometry, T(17: 1), L. *Applicable Differential Geometry.* M. Crampin, F.A.E. Pirani. London Math. Soc. Lect. Note Ser., V. 59. Cambridge U Pr, 1986, 394 pp, $34.50 (P). [ISBN: 0-521-23190-6] Essentially an introduction to the standard topics of differential geometry, including discussions of manifolds, metrics, connections, Lie groups, spinors and bundles along with over 650 exercises. The presentation is reasonably clear and it would make an adequate text for a first graduate course in differential geometry. Despite the title of the book, however, this is not a text in which the ideas of differential geometry are developed within the context of applications. AM

Differential Geometry, T(17-18: 1), S. *Lectures on Complex Analytic Manifolds.* L. Schwartz. Springer-Verlag, 1986, iv + 182 pp, $12.60 (P). [ISBN: 0-387-12877-8] Reissue of Volume 4 of Tata lecture series, first published in 1955. Twenty-five lectures, based on notes by M.S. Narasimhan, in informal style. Basics of complex variables, differential geometry comprise necessary background. Topics include differential forms and operators, de Rham's theorem, Kählerian manifolds. No frills typesetting, but a good buy. GG

Differential Geometry, P. *Lecture Notes in Mathematics-1244: Manifolds with Cusps of Rank One.* Werner Müller. Springer-Verlag, 1987, xi + 158 pp, $18 (P). [ISBN: 0-387-17696-9] Given an elliptic differential operator on an arbitrary globally symmetric space with a homogeneous vector bundle, how does one find an explicit formula for the L^2-index? Read this and find out. MR

Topology, T(17: 1), S, P. *Topological Graph Theory.* Jonathan L. Gross, Thomas W. Tucker. Ser. in Discrete Math. & Optimiz. Wiley, 1987, xv + 351 pp, $59.95. [ISBN: 0-471-04926-3] Comprehensive, "historically sensitive" introduction to the study of embeddings of graphs in surfaces such that no two edges cross. Discussion of the solution to the Heawood conjecture (determination of the chromatic numbers of non-spherical surfaces), voltage graphs, covering spaces, genus formulas, connections to other areas of mathematics. RM

Topology, P. *The Clone of a Topological Space.* W. Taylor. Res. & Expos. in Math., V. 13. Heldermann Verlag, 1986, 91 pp, $25 (P). [ISBN: 3-88538-213-X] An intriguing paper which explores how properties of a topological space A are reflected by its clone $C(A)$, the set of all continuous maps from the product of n copies of A to A where n ranges over all integers. Basic properties of A (connectedness, fixed point properties, dimension) correspond to algebraic or first order properties of $C(A)$; the most interesting results are obtained for homotopy properties. The author also shows that the clone of A can be interpreted as the set of continuous functions from A to R. LW

Operations Research, S(17-18), P. *Vehicle Routing with Time-Window Constraints: Algorithmic Solutions.* Ed: Bruce L. Golden, Arjang A. Assad. American Sciences Pr, 1986, 177 pp, $49.75 (P). [ISBN: 0-935950-15-X] Papers on vehicle routing; also appeared as the *American Journal of Mathematics and Management Sciences,* V. 6, No. 3-4, 1986. RM

Operations Research, P. *Production-inventory Control Models: Approximations and Algorithms.* A.G. de Kok. CWI Tract, V. 30. Math Centrum, 1987, iii + 214 pp, Dfl. 33 (P). [ISBN: 90-6196-310-9] From the author's introduction: "This monograph concerns the probabilistic analysis of a variety of one-product production-inventory models in which the control problem is to coordinate the production rate with the inventory level in order to cope with random fluctuations in demand." AO

Operations Research, P. *Stochastic Games with Finite State and Action Spaces.* O.J. Vrieze. CWI Tract, V. 33. Math Centrum, 1987, 221 pp, Dfl. 34.20 (P). [ISBN: 90-6196-313-3] A slightly revised version of the author's thesis. Analyzes discounted and undiscounted two-person zero sum stochastic games. Presents both structural results and algorithms. AO

Optimization, P. *Lecture Notes in Computer Science-268: Constrained Global Optimization: Algo-*

rithms and Applications. P.M. Pardalos, J.B. Rosen. Springer-Verlag, 1987, vii + 143 pp, $15.40 (P). [ISBN: 0-387-18095-8] A summary of recent work on deterministic methods (enumerative techniques, branch and bound, cutting planes, etc.) for finding global solutions of constrained optimization problems. AO

Optimization, P. *Contributions to Modern Calculus of Variations.* Ed: Lamberto Cesari. Wiley, 1986, 231 pp, $39.95 (P). [ISBN: 0-470-20378-1] Papers of fourteen authors contributed to the symposium honoring Leonida Tonelli held in Bologna on May 13-14, 1985. Includes several survey articles (historical and topical) as well as some current results. LW

Optimization, P. *Optimization Models Using Fuzzy Sets and Possibility Theory.* Ed: J. Kacprzyk, S.A. Orlovski. Theory & Decision Lib., Ser. B. Kluwer Academic, 1987, xii + 462 pp, $99. [ISBN: 90-277-2492-X] A collection of papers surveying the use of fuzzy sets and possibility theory in optimization and decision theory problems. Includes an introductory section for the non-expert. AO

Optimization, P. *Lecture Notes in Control and Information Sciences-82: Analysis and Algorithms of Optimization Problems.* Ed: K. Malanowski, K. Mizukami. Springer-Verlag, 1986, viii + 236 pp, $18.60 (P). [ISBN: 0-387-16660-2] Ten independent articles summarizing research resulting from a Japanese-Polish cooperative program entitled "Numerical Methods of Optimization and Game Theory." Most of the papers are concerned with the modeling and analysis of distributed parameter systems. AO

Dynamical Systems, P. *Cell-to-Cell Mapping: A Method of Global Analysis for Nonlinear Systems.* C.S. Hsu. Appl. Math. Sci., V. 64. Springer-Verlag, 1987, xii + 352 pp, $49.80. [ISBN: 0-387-96520-3] Introduction to the field, with many examples and 125 illustrations. Point mapping; simple cell mapping and applications; generalized cell mapping; an iterative method of global analysis. Many references. DFA

Dynamical Systems, P. *Oscillation, Bifurcation and Chaos.* Ed: F.V. Atkinson, W.F. Langford, A.B. Mingarelli. Canadian Math. Soc. Conf. Proc., V. 8. AMS, 1986, xv + 711 pp, $75 (P). [ISBN: 0-8218-6013-5] A collection of 92 lectures delivered at the 1986 Annual Seminar of the Canadian Mathematical Society. The lectures from the first week of the Conference focus on Sturmian theory, oscillation theory for second order equations, and modern generalizations. Those of the second week present results of current research in bifurcation theory and chaos. LW

Dynamical Systems, P. *Instabilities and Nonequilibrium Structures.* Ed: Enrique Tirapegui, Danilo Villarroel. Math. & Its Applic. Kluwer Academic, 1987, x + 337 pp, $69.50. [ISBN: 90-277-2420-2] A series of papers from a workshop in Valparaiso, Chile in December 1985 studying the behavior of physical systems close to and far away from equilibrium where instabilities appear. Papers cover both deterministic and stochastic behavior under the influence of noise. MR

Control Theory, T(18: 2), P. *Structured Hereditary Systems.* James A. Reneke, Robert E. Fennell, Roland B. Minton. Pure & Appl. Math., V. 107. Marcel Dekker, 1987, viii + 217 pp. [ISBN: 0-8247-7772-7] Title describes complex systems with algebraic relationships between operators (on appropriate function spaces) describing system components. Problems such as optimal control, parameter estimation, and decentralized control in four main classes of systems are studied. Many examples, some problems, and sparse index. MR

Control Theory, S(18), P. *Lecture Notes in Control and Information Sciences-99: Robust Stabilization Against Structured Perturbations.* S.P. Bhattacharyya. Springer-Verlag, 1987, ix + 172 pp, $25.10 (P). [ISBN: 0-387-18056-7] Addresses problems connected with stability of transfer functions of feedback control for engineering plants (e.g., aircraft, robots). Two topics serve as themes: first, how to determine largest hypersphere of stability; and secondly how to design transfer function stable under a given class of perturbations. MR

Control Theory, P. *Systems Analysis by Graphs and Matroids: Structural Solvability and Controllability.* Kazuo Murota. Algorithms & Combin., V. 3. Springer-Verlag, 1987, ix + 281 pp, $48 (P). [ISBN: 0-387-17659-4] An eloquent and thorough monograph which presents systematic procedures for structural analysis of large-scale systems. The theory is developed and then applied to existing as well as new models of dynamical systems. Emphasizes the importance of relevant physical observations for successful modeling. LW

Probability, T(15-16: 2), L. *Probability Theory for Engineers.* V.P. Chistyakov, B.A. Sevast'yanov, V.K. Zakharov. Transl. Ser. in Math. & Eng. Optimization Software, 1987, xiii + 161 pp, $30. [ISBN: 0-911575-13-8] Introduction to basic probability theory at the undergraduate level requiring only calculus as background. Uses random variables throughout. Clear explanations along with plenty of examples and problems. MR

Probability, P. *Lectures on Topics in Probability Inequalities.* M.L. Eaton. CWI Tract, V. 35. Math Centrum, 1987, v + 197 pp, Dfl. 30.40 (P). [ISBN: 90-6196-316-8] Based on lectures given by the author at the University of Amsterdam. Topics include majorization results and their extension to other group-induced orderings, log concavity, association, and the FKG inequality. Also includes a review of back-

ground material and an appendix on topics in convexity. RH

Stochastic Processes, T(18), P. *Stochastic Differential Systems: Analysis and Filtering.* V.S. Pugachev, I.N. Sinitsyn. Wiley, 1987, xx + 549 pp, \$95. [ISBN: 0-471-91243-3] Translated from the 1985 Russian edition. For specialists in applied mathematics. Linear and nonlinear systems. Optimal, suboptimal, conditionally optimal filtering and extrapolation. Exercises. Over 100 references to the literature. DFA

Stochastic Processes, P. *Lecture Notes in Mathematics-1233: Stability Problems for Stochastic Models.* Ed: V.V. Kalashnikov, B. Penkov, V.M. Zolotarev. Springer-Verlag, 1987, vi + 223 pp, \$19.40 (P). [ISBN: 0-387-17204-1] Proceedings of the ninth international seminar in Varna, Bulgaria, May 1985. Fifteen papers on approximation problems considered as stability problems and use of probability metrics on spaces of random variables. MR

Elementary Statistics, T(14-16: 1), C. *Introduction to Probability and Statistics for Engineers and Scientists.* Sheldon M. Ross. Ser. in Prob. & Math. Stat. Wiley, 1987, xv + 492 pp, \$32.95. [ISBN: 0-471-81752-X] A well-written, thorough introduction to probability and statistics. Assumes a knowledge of elementary calculus. The examples and exercises are plentiful and especially interesting. Includes a diskette of 35 programs for the use of students who have access to a personal computer. For students without such access, tables are included to enable them to solve the problems in the text. RH

Elementary Statistics, S(13-14). *Elementary Statistics for IBM PCs.* Donald W. Moffat. Prentice-Hall, 1988, xv + 279 pp, \$23.95 (P). [ISBN: 0-13-260050-1] A collection of BASIC programs together with examples of their use. Could be used to supplement an introductory textbook, but could not substitute for one. Unfortunately, the programs are not available on diskette and must be entered from listings in the book. AO

Elementary Statistics, T(13-14: 1, 2). *Introduction to Statistics, Fourth Edition.* Sol Weintraub. University Statistical Tracts (75-19 171st St., Flushing, NY 11366), 1987, xii + 403 pp, \$34.95. [ISBN: 0-931316-02-2] An introduction to statistics which students should find very readable. Calculus is not assumed. Many examples are worked out and plenty of exercises are included. The tables have been newly computed. Proofs and special topics are presented in the appendices. (*Third Edition*, TR, February 1979.) RH

Statistics, T*(13-14: 1, 2), S, L. *Statistics Today: A Comprehensive Introduction.* Donald Byrkit. Benjamin/Cummings, 1987, xxv + 939 pp, \$32.95. [ISBN: 0-8053-0740-0] Presupposes no college mathematics. The usual topics plus some explanatory data analysis, two-way analysis of variance, and multiple regression. MINITAB or SAS is used throughout. FLW

Statistics, T(16-18), S, P. *New Perspectives in Theoretical and Applied Statistics.* Ed: Madan Lal Puri, José Pérez Vilaplana, Wolfgang Wertz. Ser. in Prob. & Math. Stat. Wiley, 1987, xxiii + 544 pp, \$59.95. [ISBN: 0-471-84800-X] Papers from the Third International Meeting of Statistics held in Spain in 1986. Covers new developments in modeling, design, time series, density estimation, and many other areas. FLW

Statistics, S(15-18). *Modern Statistical Selection: Part II of Proceedings of the Conference on Statistical Ranking and Selection—Three Decades of Development.* Ed. M. Haseeb Rizvi. Amer. J. of Math. & Manag. Sci., V. 6, Nos. 1 & 2. American Sciences Pr, 1986, 249 pp, \$49.75 (P). [ISBN: 0-935950-14-1]

Statistics, T(15-16: 1, 2), S. *Statistics for Economics, Business Administration, and the Social Sciences.* Erling B. Andersen, Niels-Erik Jensen, Nils Kousgaard. Springer-Verlag, 1987, xi + 439 pp, \$28 (P). [ISBN: 0-387-17720-5] Translation of a Danish text. Presupposes calculus and some matrix algebra. Includes chapters on multiple regression, autocorrelation, and contingency tables. FLW

Statistics, S(18), P. *Normal Approximation and Asymptotic Expansions.* R.N. Bhattacharya, R. Ranga Rao. Robert E Krieger, 1986, xiv + 291 pp, \$46.95. [ISBN: 0-89874-690-6] Corrects misprints and adds seventeen-page chapter of updates to original 1976 Wiley edition (TR, October 1976). GG

Statistics, T(13-14: 1), S. *Modern Elementary Statistics, with Theoretical Supplement and BASIC Programming.* Donald W. Zimmerman, Richard H. Williams. Ser. in Math. & Manag. Sci., V. 16. American Sciences Pr, 1986, viii + 206 pp, \$27.50 (P). [ISBN: 0-935950-16-8] Attempts to reveal some of the theory behind elementary statistics without bogging down in mathematical details or presupposing calculus. Includes an introduction to BASIC. FLW

Elementary Computer Science, T(13: 1). *Pascal and Its Applications: An Introduction to Modern Programming.* Larry Joel Goldstein. Holt, Rinehart & Winston, 1987, xi + 619 pp, \$23.50 (P). [ISBN: 0-03-009928-5] Problem-solving approach. Emphasis on structured programming and program design. Many examples and exercises. Introduces procedures early. Effective warnings about common programming pitfalls. Attractive book with an easy-to-read, often entertaining style. DFA

Elementary Computer Science, T*(13: 1), L. *Intermediate Problem Solving and Data Structures: Walls and Mirrors.* Paul Helman, Robert Veroff. Ser. in Struct. Prog. Benjamin/Cummings, 1986, xxi + 612 pp, \$32.95. [ISBN: 0-8053-8940-7] Solid

CS2 text, with enough topics to cover a variety of courses. Review of Pascal, nice treatment of recursion ("mirrors"), data abstraction ("walls"), lists, trees, tables, sorting. Nice program development, code examples (available on disk), exercises. RM

Programming, T(13: 1). *Structured VAX BASIC.* R. Hirschfelder, *et al.* Benjamin/Cummings, 1986, xxiii + 643 pp, $28.95 (P). [ISBN: 0-8053-3690-7] Standard, readable introduction to programming with VAX BASIC, with emphasis on algorithms, pseudo-code, structured design, modularity, good style, with some advanced topics. Integrated with VAX/VMS command syntax. Sample programs in the text follow the stated guidelines. RM

Programming, P. *Engineering and Scientific Computations in PASCAL.* Lawrence P. Huelsman. Comput. Sci. & Tech. Ser. Harper & Row, 1986, xiv + 285 pp, (P). [ISBN: 0-06-042994-1] How to use Pascal for finding numerical solutions of engineering problems. Develops Pascal subroutines for numerical integration, numerical solution of differential equations, computer determination of partial fraction expansions, inverse Laplace transforms, Fourier series analysis, and others. Examples from electrical engineering clearly illustrate the application of each technique. Emphasis placed on use of numerical techniques, not on the techniques themselves. Intended for microcomputers. Includes exercises. PS

Programming, T(13: 1). *Fundamental Programming with FORTRAN 77: A Science and Engineering Approach.* J. Denbigh Starkey, Rockford J. Ross. West, 1987, xviii + 876 pp, $32.75 [ISBN: 0-314-77805-5]; *Instructor's Guide,* xviii + 519 pp, (P). [ISBN: 0-314-87243-4] Introduction to programming and to Fortran. Emphasizes program design, analysis, verification. Case studies are central. Designs programs in top-down, structured fashion using pseudo-language, then implements them in Fortran. *Instructor's Guide* discusses points to emphasize, provides answers to exercises, supplies test questions. DFA

Programming, T(13-15: 1), S, P, L*. *Prolog: A Relational Language and Its Applications.* John Malpas. Prentice-Hall, 1987, x + 465 pp, $28.67 (P). [ISBN: 0-13-730805-1] An introduction to the Prolog programming language, logic programming, applications. Prerequisite: at least one programming language. Features will appeal to wide audience: optional introduction to theoretical notions, answered exercises, artificial intelligence applications, manufacturing case study, bibliography. Examples (in "core Prolog") run on a dozen different language versions. RB

Programming, S(13-14). *Advanced Turbo C.* Herbert Schildt. Osborne/McGraw-Hill, 1987, xii + 397 pp, $22.95 (P). [ISBN: 0-07-881280-1] Turbo C is the C compiler produced by Borland International for the IBM PC and PC-compatible systems. C is a systems implementation language that is widely used by computer scientists for building software packages. This text is an introduction to the Turbo C compiler which has a number of extensions and capabilities not found in other C compilers. MS

Programming, T(13: 1). *Pascal and Algorithms: An Introduction to Problem Solving.* Gregory F. Wetzel, William G. Bulgren. Comput. Sci. Ser. SRA, 1987, xviii + 570 pp, (P). [ISBN: 0-574-18630-1] Problem-solving and algorithm development have center stage; treats programming as a useful tool for implementing solutions. Contains all the material for the first course, but syntax and semantics have secondary importance. Procedures and functions come after control structures but before structured data types. Historical anecdotes. Worth a look. DFA

Programming, S(14-15). *FORTH: The Next Step.* Ron Geere. Addison-Wesley, 1986, ix + 89 pp, $16.95 (P). [ISBN: 0-201-18050-2] FORTH is a new programming language that is widely used in the control of laboratory equipment such as machine tools, telescopes, and robots. This text is an introduction to the language for people with some programming experience, but no prior knowledge of FORTH. It uses a wide range of examples to illustrate the power and capabilities of the language. MS

Programming, S(15-16). *A Guide to the SQL Standard.* C.J. Date. Addison-Wesley, 1987, xiv + 205 pp, $25.95 (P). [ISBN: 0-201-05777-8] SQL is the official standard language for use with all relational database systems. Most relational products in the marketplace use SQL as their query language. This text is a complete introduction to the language. It shows how users can form queries in SQL to retrieve arbitrary data from the database. The emphasis is on the language standard and the principles involved, rather than on any specific product or software system. MS

Programming, T??(13), S, L. *Using Turbo Prolog.* Phillip R. Robinson. Osborne McGraw-Hill, 1987, xxv + 340 pp, $19.95 (P). [ISBN: 0-07-881253-4] Fairly complete introduction to the syntax and use of Turbo Prolog for PC's, with some discussion of advanced programming techniques and applications. RM

Programming, S. *Using Turbo C.* Herbert Schildt. Osborne/McGraw-Hill, 1988, xv + 431 pp, $19.95 (P). [ISBN: 0-07-881279-8] A primer on the C language and the programming environment provided by Turbo C. For programmers without prior experience with C. AO

Languages, P. *Languages for Sensor-Based Control in Robotics.* Ed: Ulrich Rembold, Klaus Hörmann. NATO ASI Ser. F, V. 29. Springer-Verlag, 1987, ix + 625 pp, $100. [ISBN: 0-387-17665-

9] Proceedings of the NATO Advanced Research Workshop on Languages for Sensor-Based Control in Robotics held September 1-5, 1986 in Castelvecchio Pascoli, Italy. The purpose of the workshop was to define the state-of-the-art of languages for robots and make recommendations for future research. AO

Languages, P. *Lecture Notes in Computer Science-262: A Review of Ada Tasking.* Alan Burns, Andrew M. Lister, Andrew J. Wellings. Springer-Verlag, 1987, viii + 141 pp, $15.40 (P). [ISBN: 0-387-18008-7] The Ada programming language's *task* model for parallel processing has been extensively discussed and analyzed. "This book draws on the available literature to present a comprehensive review of the Ada tasking model," evaluating this approach to concurrency programming in the context of real-time and distributed systems, with reference to formal semantics and implementation. RB

Languages, P. *Lecture Notes in Computer Science-260: ANNA: A Language for Annotating Ada Programs, Reference Manual.* David C. Luckham, *et al.* Springer-Verlag, 1987, v + 143 pp, $15.40 (P). [ISBN: 0-387-17980-1] ANNA is an extension of Ada that allows the formal specification of the intended behavior of Ada programs. These annotations are machine processable so that they can be used in conjunction with other tools. AO

Algorithms, P. *Lecture Notes on Bucket Algorithms.* Luc Devroye. Progress in Comput. Sci., No. 6. Birkhauser Boston, 1986, 146 pp, $26. [ISBN: 0-8176-3328-6] Bucket algorithms are techniques for managing the storage and retrieval of data into special data structures called hash tables. This is a set of lecture notes for a course in theoretical computer science which treats the subject of bucket algorithms and its relationship to the statistical distribution of data. MS

Computer Systems, S(16-17), P. *Testing in Software Development.* Ed: Martyn A. Ould, Charles Unwin. British Comput. Soc. Mono. in Informatics. Cambridge U Pr, 1986, x + 124 pp, $16.95 (P). [ISBN: 0-521-33786-0] Guide to the management of testing of software throughout the life cycle (requirements specification, design, implemented system) from the viewpoints of the user, development manager, designers, programmers. Key ideas are that testing must be considered broadly (as informal design reviews, test case analyses, formal proofs), and must be managed throughout the process so that test results become part of the project throughout. RM

Computer Systems, P. *An Introduction to Text Processing Systems: Current Problems and Solutions.* Ed: J.J.H. Miller. Boole Pr, 1985, vii + 120 pp. [ISBN: 0-906783-51-8] Notes from 11 lectures given at an October 1985 workshop held in conjunction with the Protext II Conference in Dublin. Includes introductions to TeX, Metafont, and the Standard Generalized Markup Language (SGML). LAS

Computer Systems, P, L. *UNIX System Readings and Applications.* AT&T. Prentice-Hall, 1987, $18 each (P). *Volume I: UNIX Time-Sharing System,* xiv + 397 pp [ISBN: 0-13-938532-0]; *Volume II: The UNIX System,* xii + 324 pp. [ISBN: 0-13-939845-7] Reprints of two special issues of the *Bell System Technical Journal* (July-August 1978, October 1984) devoted to the UNIX operating system and extensions. UNIX was developed at Bell Laboratories beginning about 1970, and offers elegant, powerful design features. It is heavily used in educational institutions and increasingly in commercial settings; a considerable body of related software has evolved. These 38 articles survey the system and basic applications (*Volume I*) as well as later developments (*Volume II*), and are largely written by the authors of the system and applications themselves. The readings are generally less terse, more expository than comparable entries in the UNIX Programmer's Manual. A collection of important, often-cited background papers. RB

Computer Systems, P. *Lecture Notes in Computer Science-257: Database Machine Performance: Modeling Methodologies and Evaluation Strategies.* Ed: Francesca Cesarini, Silvio Salza. Springer-Verlag, 1987, x + 250 pp, $23.10 (P). [ISBN: 0-387-17942-9] Thorough coverage of methods for analyzing performance of database machines—machines designed specifically to deal with databases. Suggestion of performance indexes, measuring methods, and strategy for machine analysis. Includes performance analysis of existing database machines, DBMAC and VERSO. Stresses the importance of analyzing and modeling during the development stages of a database machine. PS

Computer Systems, P. *Lecture Notes in Computer Science-258 & 259: PARLE: Parallel Architectures and Languages Europe.* Ed: J.W. de Bakker, A.J. Nijman, P.C. Treleaven. Springer-Verlag, 1987, $34.60 each (P). *Volume I: Parallel Architectures,* xii + 479 pp [ISBN: 0-387-17943-7]; *Volume II: Parallel Languages,* xii + 463 pp. [ISBN: 0-387-17945-3] Proceedings of a conference on theory, application and design of parallel computer systems in Eindhoven, The Netherlands, June 1987, initiated by Project 415 of the European organization ESPRIT. 54 research papers by European authors on topics including: concurrent, object-oriented, logic and functional programming; MIMD and reduction parallel computers; process theory, design and verification of parallel systems, performance evaluation; interconnection networks, systolic arrays; VLSI, RISC architectures; applications and special purpose architectures. Includes overviews of the field (invited lectures and ESPRIT 415 reports). RB

Computer Systems, P. *Parallel Computer Vision.*

Ed: Leonard Uhr. Academic Pr, 1987, xiv + 303 pp, $29.95. [ISBN: 0-12-706958-5] A pyramid is a massively parallel multi-computer network consisting of a tree-like hierarchy of computers interconnected as arrays at each layer, such that successively higher layers contain fewer computers. These eleven research papers concern hardware architecture, algorithms and software for pyramids and variations, and applications of such systems to computer vision. RB

Computer Systems, P. *Evaluation of Multicomputers for Image Processing.* Ed: L. Uhr, *et al.* Academic Pr, 1986, xiv + 346 pp, $35. [ISBN: 0-12-706962-3] A benchmark is a computer program whose execution time serves as a measurement standard for performance comparisons between computer systems. This book, proceedings of the 1984 Tanque Verde workshop on multicomputer systems in image processing (Tucson, Arizona), focuses on benchmarking and other evaluation techniques appropriate for such systems. Seventeen research papers. RB

Computer Systems, P. *Smalltalk-80: Bits of History, Words of Advice.* Ed: Glenn Krasner. Addison-Wesley, 1983, viii + 344 pp, $25.95 (P). [ISBN: 0-201-11669-3] Smalltalk-80 is a software system representing on-going efforts at Xerox Corporation's Palo Alto Research Center (PARC) to make large amounts of computing power accessible to users. Xerox invited outside groups to implement Smalltalk-80 on a variety of hardware. This book describes the implementors' experiences. Primarily for Smalltalk-80 implementors, software engineers. RB

Computer Systems, P. *Design of VLSI Circuits Based on VENUS.* E. Hörbst, C. Müller-Schloer, H. Schwärtzel. Springer-Verlag, 1987, xii + 318 pp, $59.50. [ISBN: 0-387-17663-2] A textbook on designing VLSI circuits using the VENUS product (Siemens AG, Munich). This CAD system facilitates representation, design and verification, with accommodations for manufacture and testing. Summary of integrated circuit design methods, microelectronic technologies, layout design methods, and testing principles; practical VENUS instructions and cell libraries; outlook. RB

Computer Systems, P. *Mechanisms for Reliable Distributed Real-Time Operating Systems: The Alpha Kernel.* J. Duane Northcutt. Perspect. in Comput., V. 16. Academic Pr, 1987, xiv + 245 pp, $25. [ISBN: 0-12-521690-4] A research monograph describing a largely implemented operating system kernel (Alpha kernel) which integrates updated techniques (including decentralized management of global resources, object orientation, exclusion of policy from kernel, deadlines) in a modular distributed real-time system; created as a vehicle for the study of issues in decentralized operating systems. RB

Computer Systems, P. *Lecture Notes in Computer Science-272: Future Parallel Computers.* Ed: P. Treleaven, M. Vanneschi. Springer-Verlag, 1987, v + 492 pp, $34.60 (P). [ISBN: 0-387-18203-9] A collection of topics following the structure of a June 1986 Advanced Course held in Pisa. Covers main classes of parallel computers, illustrates these classes by examining important parallel systems being developed, and presents topics that influence all classes of parallel computers. PS

Computer Systems, T(14-16), S, P*. *A Guide to INGRES.* C.J. Date. Addison-Wesley, 1987, xiii + 385 pp, $32.95. [ISBN: 0-201-06006-X] A detailed description of the INGRES relational database management system, intended for end-users and application programmers in data processing, database professions, including students and teachers of data processing. Thorough coverage of user-oriented material, e.g., SQL query language; sketchy on details of interest only to system programmers or machine operators. Highly readable, highly informative. RB

Computer Systems, P. *Computer Aided Text Processing—An Introduction.* Ed: J.J.H. Miller. Boole Pr, 1986, vii + 69 pp. [ISBN: 0-906783-57-7] Lecture notes from a short course on TEX, LaTeX, Postscript, and SGML (Standard Generalized Markup Language) held in conjunction with the October 1986 Protext III Conference in Dublin. LAS

Computer Systems, P. *Lecture Notes in Computer Science-253: WOPPLOT 86: Parallel Processing: Logic, Organization, and Technology.* Ed: J.D. Becker, I. Eisele. Springer-Verlag, 1987, vi + 226 pp, $20.60 (P). [ISBN: 0-387-18022-2] Proceedings of a workshop at Neubiberg, FRG (July 1986) featuring reports on current work and future prospects of parallel processing. Silicon vs. molecular electronics technology; organizational structure of parallel processors, including pyramid architectures and MIMD (multiple instruction/multiple data) machines; applications to physics and image processing. RB

Computer Graphics, T(13), P*. *Device-Independent Graphics with Examples from IBM Personal Computers.* Robert F. Sproull, W.R. Sutherland, Michael K. Ullner. McGraw-Hill, 1985, xii + 546 pp, $35.95 (P). [ISBN: 0-07-060504-1] The Graphical Kernel System (GKS) is a device-independent graphics subroutine package that is an approved international standard. This book is an introduction to the design and programming of interactive graphics applications. It includes an introduction to GKS and examples of programs using GKS. Also includes a discussion of device-independent graphics software and standards. JAS

Computer Graphics, P, L. *Introduction to the Graphical Kernel System (GKS), Second Edition.* F.R.A. Hopgood, *et al.* APIC Stud. in Data Processing, No. 28. Academic Pr, 1986, xii + 250 pp, $23.95 (P). [ISBN: 0-12-355571-X] A descriptive

introduction with examples to GKS, the first (programming language-independent) graphics system to be endorsed by the International Standards Organization (ISO, 1985), by the authors of the ISO document. Essential concepts; features for the specialist. Intended for the application programmer with at least rudimentary knowledge of computer graphics. RB

Computer Graphics, P, L., *Computer Graphics Programming: GKS—The Graphics Standard, Second, Revised and Enlarged Edition.* G. Enderle, K. Kansy, G. Pfaff. Symb. Computat. Ser. Springer-Verlag, 1987, xxiii + 651 pp, $59. [ISBN: 0-387-16317-4] A complete reference on the GKS standard. This edition reflects changes made to the standard after 1983 and adds a new section on the three-dimensional extension of GKS. (*First Edition,* TR, November 1984.) AO

Theory of Computation, P. *Lecture Notes in Computer Science-256: Rewriting Techniques and Applications.* Ed: Pierre Lescanne. Springer-Verlag, 1987, vi + 285 pp, $23.10 (P). [ISBN: 0-387-17220-3] Proceedings of second conference on Rewriting Techniques and Applications, Bordeaux, May 1987. Twenty-two research papers and one invited lecture on rewriting, a part of formal language theory with computer science applications involving substitution, pattern matching, formal logic, combinatorial group theory, etc. Implementation; Knuth-Bendix theorem; theoretical aspects; unification; efficiency. RB

Theory of Computation, T, P, L. *Computability.* Klaus Weihrauch. EATCS Mono. on Theor. Comput. Sci., V. 9. Springer-Verlag, 1987, x + 517 pp, $59.50. [ISBN: 0-387-13721-1] A textbook on recursion theory, presenting a framework of foundations rather than a comprehensive survey of all branches of the field. Part 1: Introduction, computability on natural numbers, words; written at undergraduate level. Part 2: Recursion theory of natural numbers, related topics. Part 3: Computability on uncountable sets (Type 2 computability), applications. Each part largely self-contained; hundreds of exercises. RB

Artificial Intelligence, T??, S??, P? *Artificial Intelligence Using C.* Herbert Schildt. Osborne McGraw-Hill, 1987, xi + 412 pp, $21.95 (P). [ISBN: 0-07-881255-0] A survey of some programming issues related to artificial intelligence, geared toward a popular audience. Considers expert systems, natural language processing, computer vision, pattern recognition, robotics, machine learning, etc. Presented as "how-to" manual; necessarily superficial; very debatable opinions expressed as fact, e.g., "[t]here are only vague historical reasons why special artificial intelligence languages [LISP, Prolog] were invented in the first place." 63-page review of C. RB

Computer Science, S(17), P. *Lecture Notes in Computer Science-245: Lectures on the Complexity of Bilinear Problems.* H.F. de Groote. Springer-Verlag, 1987, 135 pp, $15 (P). [ISBN: 0-387-17205-X] Based on the author's 1982 lectures given at the University of Zurich. The objective is to present the complexity theory of bilinear problems in a unified and coordinate-free form. The emphasis is on those problems which are defined by multiplication in associative algebras. CEC

Computer Science, P. *Lecture Notes in Computer Science-261: Translating Relational Queries into Iterative Programs.* Johann Christoph Freytag. Springer-Verlag, 1987, xi + 131 pp, $15.40 (P). [ISBN: 0-387-18000-1] This dissertation investigates the automatic translation of set-oriented relational database query specifications into iterative programs, using functional programming and program transformation techniques. The author presents two algorithms for generating the iterative programs, considers efficient programs for evaluating aggregate functions (e.g., average), and reports on Lisp (T) as a database system implementation language. RB

Computer Science, P, L. *Advances in Computers, Volume 26.* Ed: Marshall C. Yovits. Academic Pr, 1987, xi + 476 pp, $65. [ISBN: 0-12-012126-3] This annual publication consists of survey articles on selected topics by various authors, written from a relatively leisurely perspective. Computer support of "fuzzy" human reasoning, unary (vs. binary) representation in computers, parallel algorithms, multistage interconnection networks, fault-tolerant computing, VLSI validation and testing, software testing concepts, development of reliable distributed software systems. RB

Computer Science, P. *Concurrency and Nets: Advances in Petri Nets.* Ed: K. Voss, H.J. Genrich, G. Rozenberg. Springer-Verlag, 1987, xiii + 622 pp, $49. [ISBN: 0-387-18057-5] Addresses and papers in honor of C.A. Petri's 60th birthday, plus many contributed papers on current issues and applications of nets. RM

Applications, P. *Proceedings of the First International Conference on Industrial and Applied Mathematics (ICIAM 87).* Ed: A.H.P. van der Burgh, R.M.M. Mattheij. CWI Tract, V. 36. Math Centrum, 1987, 433 pp, Dfl. 56.80 (P). [ISBN: 90-6196-318-4] Contains the papers presented at ICIAM 87 by researchers in the Netherlands. The 29 papers in this volume cover a wide range of topics in applied and industrial mathematics. AO

Applications, P. *Protext III.* Ed: J.J.H. Miller. Boole Pr, 1987, vii + 169 pp. [ISBN: 0-906783-55-0] Proceedings of the October 1986 third international conference on text processing held in Dublin. Focus on font design and user interface in computer typesetting systems. LAS

Applications (Biological Science), P. *Lecture Notes in Biomathematics-71: Mathematical Topics in Population Biology, Morphogenesis and Neurosciences.* Ed: E. Teramoto, M. Yamaguti. Springer-Verlag, 1987, ix + 348 pp, $37.40 (P). [ISBN: 0-387-17875-9] Proceedings of a 1985 international symposium in Kyoto, Japan. Papers on mathematical ecology, population biology (structure, dispersal, and evolution), theory of patterns and morphogenesis, neuroscience, and physiology. RM

Applications (Biological Science), P. *Lecture Notes in Biomathematics-70: Stochastic Methods in Biology.* Ed: M. Kimura, G. Kallianpur, T. Hida. Springer-Verlag, 1987, vi + 229 pp, $22.80 (P). [ISBN: 0-387-17648-9] Proceedings of a 1985 workshop in Nagoya, Japan for biologists and mathematicians. Topics of papers include population genetics, neurophysiology, fluctuations in living cells, mathematical treatments of epidemiology, population dynamics, stochastic differential equations. RM

Applications (Communication Theory), T(16: 1), P. *Digital Transmission Theory.* Sergio Benedetto, Ezio Biglieri, Valentino Castellani. Prentice-Hall, 1987, xvi + 639 pp, $53.33. [ISBN: 0-13-214313-5] Treatment of digital communication theory (introduction to mathematical techniques, information theory, modulation schemes, Gaussian channels, adaptive receivers, encoding and reliability, non-linear channels) with emphasis on engineering applications, bandwidth, power, and complexity of systems as major design trade-offs. RM

Applications (Economics), P. *Mathematical Economics.* Kelvin Lancaster. Dover, 1987, xiii + 411 pp, $9.95 (P). [ISBN: 0-486-65391-9] Republication of the work originally published by Macmillan in 1968 (TR, November 1968). PS

Applications (Electrical Engineering), T(16-17: 1). *Mathematical Modeling and Digital Simulation for Engineers and Scientists, Second Edition.* Jon M. Smith. Wiley, 1987, xv + 430 pp, $36.95. [ISBN: 0-471-08599-5] Major additions in this edition are new sections on numerical methods for simulating random processes and simulator verification. New material has also been added throughout to better explain the use of numerical techniques. (*First Edition,* TR, October 1978.) AO

Applications (Electrical Engineering), P. *Operator Theory, Analytic Functions, Matrices, and Electrical Engineering.* J. William Helton, *et al.* CBMS Reg. Conf. Ser. in Math., No. 68. AMS, 1987, xiv + 134 pp, $19 (P). [ISBN: 0-8218-0718-8] Expansion of ten lectures given at a regional conference in Lincoln, Nebraska. Primarily useful as a list of theorems and references. GG

Applications (Engineering), P. *Numerical Methods in Transient and Coupled Problems.* Ed: R.W. Lewis, *et al.* Wiley, 1987, xvi + 350 pp, $95. [ISBN: 0-471-91200-X] Of interest to reseasrchers in civil engineering. LC

Applications (Engineering), T(16-17: 1), P, L. *Solving Polynomial Systems Using Continuation for Engineering and Scientific Problems.* Alexander Morgan. Prentice-Hall, 1987, xiii + 546 pp, $49.95. [ISBN: 0-13-822313-0] A continuation method (embedding or homotopy method) for solving small systems of polynomial equations which requires no initial guess and finds all solutions. Provides the necessary differential topology. Includes exercises; many suggestions for experiments and projects; case studies from real applications. Pleasant, conversational style. Contains Fortran source code (available on diskettes) for implementation. DFA

Applications (Fluid Mechanics), S(18), P. *Turbulence and Random Processes in Fluid Mechanics.* M.T. Landahl, E. Mollo-Christensen. Cambridge U Pr, 1986, vi + 154 pp, $17.95 (P). [ISBN: 0-521-34687-8] Masters-level monograph on subject intended to provide background necessary for reading the literature and beginning research. Assumes readers from a variety of backgrounds. Covers stability, wave motion, and coherent structure. MR

Applications (Information Theory), P, L. *Analysis of Fuzzy Information, Volume I-III.* Ed: James C. Bezdek. CRC Pr, 1987, $450 set. *V. I: Mathematics and Logic,* 272 pp [ISBN: 0-8493-6296-2]; *V. II: Artificial Intelligence and Decision Systems,* 251 pp [ISBN: 0-8493-6297-0]; *V. III: Applications in Engineering and Science,* 296 pp. [ISBN: 0-8493-6298-9] Fifty chapters based on papers presented at a July 1984 conference, including ten lengthy survey papers. A comprehensive survey of Lofti Zadeh's innovative concept of a fuzzy set based on imprecise information, revealing extensive applications to analysis, logic, expert systems, database systems, pattern recognition, and operations research. LAS

Applications (Management), P. *Decision Support Systems: Theory and Application.* Ed: Clyde W. Holsapple, Andrew B. Whinston. NATO ASI Ser. F, V. 31. Springer-Verlag, 1987, x + 500 pp, $90. [ISBN: 0-387-17774-4] Proceedings of NATO Advanced Study Institute (1985) on current DSS views and future trends. Recurring emphasis of artificial intelligence explored throughout. Papers cover theory (paradoxes, relational theory of model management, conceptual modelling, artificial intelligence and knowledge-based DSS), and applications (office information systems, accounting and auditing, strategic planning, and management). RM

Applications (Physics), T(18: 1), S, P. *Hamiltonian Methods in the Theory of Solitons.* L.D. Faddeev, L.A. Takhtajan. Springer-Verlag, 1987, ix + 592 pp, $110. [ISBN: 0-387-15579-1] The first of a two-volume series, this book is devoted to the classical rather than the quantum analysis of the non-

linear Schrödinger equation, with emphasis on the Hamiltonian structure. The authors have attempted to make this text sufficiently self-contained to serve as an independent introduction. MU

Applications (Physics), T(18: 2), P. *Gauge Field Theories.* Stefan Pokorski. Mono. on Math. Phys. Cambridge U Pr, 1987, xiii + 394 pp, $89.50. [ISBN: 0-521-26537-1] Gauge theory in electrodynamics studies the invariance of the Lagrangian density of fields with respect to local phase transformations. This book provides basic tools in theory of fundamental interactions requiring only a background in quantum field theory. Topics covered include basics of perturbation theory and areas of research interest (e.g., chiral symmetry). Includes many exercises. MR

Applications (Physics), T, S, P, L. *Manifolds and Mechanics.* Arthur Jones, Alistair Gray, Robert Hutton. Australian Math. Soc. Lect. Ser., V. 2. Cambridge U Pr, 1987, 166 pp, $39.50; $13.95 (P). [ISBN: 0-521-33375-X; 0-521-33650-3] A textbook aiming to make basic ideas about differentiable manifolds readily available to applied mathematicians and theoretical physicists while exhibiting applications (to Lagrangian mechanics) to the mathematician. Prerequisite: advanced calculus, e.g., Spivak's *Calculus on Manifolds.* Fundamental notions; a geometric derivation of Lagrange's equations; flows, spherical pendulum, rigid bodies. Numerous exercises. RB

Applications (Physics), P. *Percolation Theory and Ergodic Theory of Infinite Particle Systems.* Ed: Harry Kesten. IMA, V. 8. Springer-Verlag, 1987, xi + 323 pp, $34. [ISBN: 0-387-96537-8] Proceedings of a workshop at the Institute for Mathematics and its Applications, mostly concerned with random processes (such as self-avoiding walks) on lattices. Percolation theory and interacting particle systems are subfields of probability, related to statistical mechanics. BC

Applications (Physics), P, L. *Einstein's Dream: The Search for a Unified Theory of the Universe.* Barry Parker. Plenum Pr, 1986, ix + 287 pp, $18.95. [ISBN: 0-306-42343-X] A presentation for a popular audience of background and concepts in search of a "grand unified theory (GUT)" in physics. General relativity, phenomena including white dwarf stars and black holes, origins of the universe, particle physics and field interactions, current GUTs are surveyed in a historical context; "mathematics has been completely avoided in this book ...." RB

Applications (Physics), P. *Theory and Applications of Liquid Crystals.* Ed: J.L. Ericksen, D. Kinderlehrer. IMA, V. 5. Springer-Verlag, 1987, xii + 353 pp, $37. [ISBN: 0-387-96546-7] Thirteen papers, largely the proceedings of a workshop at the Institute for Mathematics and Its Applications, Minnesota, January 1985. Almost no mathematicians were engaged in liquid crystal research at that time. The workshop succeeded in stimulating subsequent work: three of the papers report on investigations undertaken after the workshop took place. RB

Applications (Physics), T(15-17: 1), L. *One-dimensional Stefan Problems: An Introduction.* James M. Hill. Mono. & Surv. in Pure & Appl. Math., V. 31. Longman Scientific & Tech (US Distr: Wiley), 1987, xviii + 204 pp, $106. [ISBN: 0-470-20388-9] Mainly studies semi-analytical solution methods for those heat-diffusion moving-boundary problems, pseudo-steady-slate approximations, large Stefan number expansion, integral formulation, upper and lower bounds for the motion, formal series solutions, integral iteration techniques; Goodman's and Megerlin's methods and an iterative series solution procedure. For practitioner as well as student. DFA

Applications (Physics), S*(13), L. *Realistic Applications in Mechanics.* The Spode Group. Oxford U Pr, 1986, viii + 76 pp, $19.95 (P). [ISBN: 0-19-853654-2] A series of explicit case studies in (calculus-based) mechanics applied particularly to sports (e.g., the "miracle" high jump in the 1968 Mexico City Olympics) and driving (designing automobile brakes). Intended to help motivate 16-19 year old vocational students in Great Britain. An excellent source of rich, supplementary problems. LAS

Applications (Physics), P. *Proceedings Seminar 1983-1985: Mathematical Structures in Field Theories, V. 2.* M.J. Bergvelt, G.M. Tuynman, A.P.E. ten Kroode. CWI Syllabus, V. 13. Math Centrum, 1987, iii + 206 pp, Dfl. 13.60 (P). [ISBN: 90-6196-317-6] Consists of three papers: "An introduction to classical mechanics and symplectic geometry," by G.M. Tuynman; "The Hamiltonian structure of Yang-Mills theories," by M.J. Bergvelt; and "Geometrical description of the Toda lattice," by A.P.E. ten Kroode. MU

Reviewers

RJA: Richard J. Allen, St. Olaf; DFA: David F. Appleyard, Carleton; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; LC: Laura Chihara, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; SG: Steven Galovich, Carleton; GG: George Gilbert, St. Olaf; BH: Bruce Hanson, St. Olaf; RH: Rhonda Hatcher, St. Olaf; RK: Roger Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; AM: Alan Magnuson, St. Olaf; SM: Steve McKelvey, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; MR: Matthew Richey, St. Olaf; AWR: A. Wayne Roberts, Macalester; MS: Michael Schneider, Macalester; PS: Paul Schoening, St. Olaf; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; LW: Lynne Walling, St. Olaf; FLW: Frank L. Wolf, Carleton; MZ: Michael Zielinski, St. Olaf.

List of Referees for the Monthly

Alexander Abian, Janos Aczel, Adebisi Agboola, John F. Ahner, Michael Albertson, Felix R. Albrecht, David J. Aldous, J. Ralph Alexander, Frank Alken, Charles Allen, Richard Allen, Ronald Alter, Richard Anderson, David Andrew, George E. Andrews, Larry C. Andrews, Tom Apostol, Kenneth Appell, Michael Aschbacher, Robert Ash, Richard Askey, Gennady Bachman, Richard Bagby, Craig Bailey, I. N. Baker, Mitchell Baker, Tom Banchoff, Steven Bank, Edward Barbeau, Ruth Bari, Charles W. Barnes, Robert Bartle, Bill Barker, Richard Bassein, Paul Bateman, Willard Baxter, Lowell W. Beineke, Janet Beissinger, David Bellamy, J. J. Benedetto, George Benke, Don Bentley, Sterling K. Berberian, I. David Berg, Deb Bergstrand, Kenneth Berman, Bruce Berndt, Stephen Bernfeld, S. P. Bhattacharyya, Ted Bick, Thomas Bickel, Norman Biggs, Richard Bishop, Gary Bloom, Mary Boas, Ralph Boas, Ken Bogart, Paul Bolgiano, Jonathan Borwein, Peter Borwein, Michael Botsko, William E. Boyce, Fred Brauer, Joel L. Brenner, Robert Breusch, William Briggs, Roger Brockett, Foster Brooks, Ron Brown, T. C. Brown, Richard Brualdi, Barry W. Brunson, J. A. Brzozowski, James Buchanan, Edwin Buchman, Creighton Buck, Richard Bumby, John W. Bunce, Richard Burden, Frank Burk, Donald Burkholder, Richard D. Byrd, Phil Byrne, Jim Callahan, Douglas M. Campbell, Stephen L. Campbell, Rodney Canfield, Frank Cannonito, G. T. Cargo, B. C. Carlson, George F. Carrier, Phyllis Cassidy, Beda Chan, Lindsay Childs, Jal Choksi, Fan R. K. Chung, Joseph A. Cima, Barry Cipra, George Cobb, Jane Coffee, John Coffey, David Cohen, Paul M. Cohn, Robert Connelly, John B. Conway, Anne Cook, Pamela Cook, Roger Lee Cooke, Curtis Cooper, Clifton Corzatt, C. C. Cowen, David Cox, Thomas Craven, James P. Crawford, Philip S. Crooke III, James Cross, Richard Crowell, Helen F. Cullen, Frederic Cunningham, Robert Currier, Everett Dade, John D'Angelo, Richard B. Darst, Chandler David, Philip Davis, Mahlon Day, Carl DeBoor, Frank DeMeyer, John Dettman, Frank Deutsch, James Deveney, Carl DeVito, Beverly Diamond, Joe Diestel, Robert A. DiPaolo, Peter Grant Dixon, David Dobbs, Avron Douglis, Peter J. Downey, Vladmir Drobot, Daniel Drucker, Roy Dubisch, Underwood Dudley, Charles F. Dunkl, Alan Durfee, William Durfee, John Dyer-Bennett, Jerome Eastham, B. C. Eaves, Gary Ebert, Michael Edelstein, Hugh M. Edgar, Bruce Edwards, Gove Effinger, Roger Eggleton, Ernest G. Enns, David Erbach, Paul Erdos, Michael Jon Evans, Ronald Evans, Howard Eves, Kenneth J. Falconer, Hanafi K. Farahat, Burton Fein, Jacob Feldman, Helaman Ferguson, Pamela Ferguson, James Fife, Michael A. Filaseta, Stephen D. Fisher, Leopold Flatto, David J. Foulis, Gilles Fournier, Gary Fowler, John B. Fraleigh, D. Freedman, Gunther Frei, Orrin Frink, Daniel Frohardt, Wolfgang Fuchs, Joseph Gallian, Fred Galvin, Martin Gardner, Richard Gardner, Bernard Gelbaum, Susan Geller, I. C. Gentry, Ira Gessel, Robert Gethner, Marlene Gewand, Peter Giblin, Jane Gilman, Anthony Gioia, Steven Givant, M. Lawrence Glasser, Leon Glesser, Herman Gluck, Don Goldberg, Sandy Grabiner, James Graham-Eagle, Richard Grassl, Alfred Gray, Leonard J. Gray, Daniel Grayson, Gary Greenfield, Philip S. Griffin, Phillip Griffith, J. Griggs, Emil Grosswald, Stan Gudder, Victor W. Guillemin, Peter Hagis, Heini Halberstam, Edwin Halfar, David Hallenbeck, Mary-Elizabeth Hamstrom, Phil Hanlin, Ferdinand Haring, Robert H. Haring-Smith, Robert E Hartwig, Kevin J. Hastings, Joseph Hemmeter, Jim Henle, Ralph Henstock, Eugene Herman, Edwin Hewitt, Thomas Hill, Peter Hilton, Larry P. Hoehn, R. A. Honsberger, Roger A. Horn, John Horvath, John Hosack, George Hsiao, Wu-teh Hsiang, Bruce Hughes, John Hughes, Rhonda Hughes, Tony Hughes, James E. Humphreys, Martin I. Isaacs, Nathaniel Jacobson, Robert C. James, Gerald Janusz, Charles Jepsen, Richard Jerrard, Carl Jockusch, Jr., Jerry Johnson, Norman Johnson, Wells Johnson, Anju Joglekar, Charles A. Jones, William R. Jones, Dan Kalman, Ravi Kannan, Mark Kannowski, Wilfred Kaplan, Irving Kaplansky, Robert Kaufman, Nicholas Kazarinoff, Jerry Kazdan, James Keesling, Leroy Kelly, Franklin Kemp, Judy Kennedy, Mark Kidwell, Allen Kirch, Kit R. Kittappa, Murray S. Klamkin, Victor Klee, Alex Kleiner, Joseph Klerlein, Dennis Kletzing, Fred Knight, Neal Koblitz, Mark Krusemeyer, Ray Kunze, John Labute, Martin D. Landau, Joachim Lazbek, Peter Lancaster, Vincent LaRiccia, Loren Larson, Ronald J. Leach, Solomon Leader, Derrick Lehmer, Y. J. Leung, Ronnie Levy, Leonard Lewin, D. J. Lewis, H. C. Li, Tien-Yien Li, Albert Livingston, Peter Loeb, Calvin Long, Robert Long, Judith Longyear, Warren Loud, Anne L. Luddington, Erwin Lutwak, Roger C. Lyndon, Charles MacCluer, Thomas Mahar, Reza Malek-Madani, Joseph Malkevitch, Walter Mallory, Colin L. Mallows, Ernie Manes, Lois Mansfield, Daniel Marcus, Morris Marden, Harold Marin, Jerrold E. Marsden, Mario Martelli, Peter H. Maserick, Paul Massell, Leon

Mattics, Daniel Mauldin, Jim Mauldon, Steven Maurer, Louis F. McAuley, Roger McCann, John McCleary, O. Carruth McGehee, Richard McGovern, John McKay, Glen Meeden, Kathy Merrill, Burnett Meyer, Mark Meyerson, George Miel, David Minda, Courtney Moen, R. A. Mollin, George Monk, Hugh Montgomery, Susan Montgomery, Emily Moore, Tom Moore, John Morrison, W. O. J. Moser, Martin Muldoon, David Muller, Harry Mullikin, James Munkres, J. Mycielski, Richard Nau, Roger Nelsen, William A. Newcomb, Morris Newman, Albert Nijenhuis, Ivan Niven, Katsumi Nomizu, Peter Norman, Karl Norton, George O'Brien, Andrew M. Odlyzko, Ingram Olkin, Michael Ong, James M. Ortega, Don O'Shea, John Oxtoby, P. P. Palfy, Francesco Palmieri, Frank Papp, Beresford Parlett, N. Tenney Peck, Jaak Peetre, Don Pelletier, Joan Pelletier, David Penney, Walter Philipp, Henry Pollak, Carl Pomerance, Horacio Porta, Stephen Portnoy, Malcolm Pownall, Donald A. Quarles, Jr., P. Rabinowitz, John R. Rabung, Q. I. Rahman, Anthony Ralston, Daniel Rawsthorne, Maxwell O. Reade, Don Redmond, David Reiner, Irving Reiner, Edward Reingold, Bruce Reznick, Billy Rhoades, Norman Rice, Ian Richards, Bruce Richter, Theodore J. Rivlin, Wayne Roberts, Derek Robinson, Donald Robinson, R. T. Rockafellar, David Roeder, C. Ambrose Rogers, Alexander Rosa, David A. Rose, Israel H. Rose, Michael Rosen, Kenneth Ross, Joseph Rotman, Halsey L. Royden, Walter Rudin, Bruce Sagan, Hans Sagan, Chin-Han Sah, Eugene Salamin, David Saltman, Tom Sanders, Doris Schattschneider, Albert Schild, Frank W. Schmidt, Georg Schmidt, Hans Schneider, Todd Schneider, Bert Schreiber, Sy Schuster, Rolph Schwarzenberger, Michael O. Searcoid, John Selfridge, Marjorie Senechal, Caroline Series, Daniel Shanks, Harold Shapiro, Jay Shapiro, Joel H. Shapiro, Dottie Sherling, Allen L. Shields, R. M. Shortt, Alan Shuchat, Stuart Sidney, Joseph H. Silverman, Evelyn M. Silvia, Rodica Simion, David A. Singer, Patricia Sipe, Neil Sloane, David Smith, Craig Smorynski, J. C. Snader, Laurie Snell, Murray R. Spiegel, Claudia Spiro, David Spring, Bhama Srinivasan, Dennis Stanton, Ralph Stanton, B. M. Stewart, Gilbert Strang, Karl Stromberg, A. H. Stone, M. V. Subbarao, William D. Sudderth, Frank Tall, Lajos Takacs, Leon Tatevossian, Alan Taylor, Howard Taylor, John A. Thorpe, Joanne Trimble, Hale Trotter, George A. Tsintsifas, Stephen Ullom, Abraham Ungar, Jeffery D. Vaaler, Charles Van Loan, Richard Varga, Richard Vitale, Andrew Vogt, W. Vollmerhaus, Rudolf Vyborny, Helen Wang, Elbert A. Walker, Henry Walker, Gary L. Walls, William Wardlaw, William Waterhouse, William Watkins, Bob Weaver, Elliott Weinberg, Douglas West, Thelma West, John Wetzel, Dennis White, Albert Whiteman, Susan Whitesides, Roger Wiegand, Sylvia Wiegand, Albert Wilansky, John B. Wilker, Richard Williamson, Gareth Williams, Hugh C. Williams, Kenneth Williams, Els Withers, Frank Wolf, John W. Wrench, Shinji Yamashita, David Yetter, James A. Yorke, Joseph Zaks, Lawrence Zalcman, Daniel Zelinsky, Paul Zorn, William Zwicker.

Shenk

Calculus and Analytic Geometry, 4/e

In single- and split-volume editions.

© 1988, 996 pages, hardbound

With Overhead Transparencies, Student Handbook, Software Toolkit, Instructor's Guide and Solutions Manual, and Computer-Assisted Testing System (CATS).

Schiller / Wurster

College Algebra and Trigonometry:

Basics Through Precalculus

© 1988, 704 pages, hardbound#

College Algebra

© 1988, 528 pages, hardbound#

#With Overhead Transparencies, Study and Solutions Manual, Instructor's Manual, and Computer-Assisted Testing System (CATS).

Steffensen / Johnson

Essential Mathematics for College Students:

Prealgebra, Algebra, and Geometry

© 1988, 768 pages, softbound

With Student Solutions Manual, Instructor's Guide, and Computer-Assisted Testing System (CATS).

Peter / Welch

Algebra for College Students

© 1988, 592 pages, hardbound

With Student Solutions Guide, Instructor's Guide, and Computer-Assisted Testing System (CATS).

Cullen

Linear Algebra with Applications

© 1988, 480 pages, hardbound

With Instructor's Guide (including Software).

Computer-Assisted Testing System (CATS)

We continue to be a leader in developing and publishing innovative software to better serve you and your students. CATS, for Apple II and IBM-PC microcomputers, is a self-contained testing system with complete editing, integrated graphics, and printout capabilities. Efficient and user-friendly, it allows you to create tests to fit your individual teaching style. CATS is menu-driven and requires no prior computer experience.

For further information write
Meredith Hellestrae,
Department SA-AMM
1900 East Lake Avenue
Glenview, Illinois 60025

U.S. Postal Service

STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION
Required by 39 U.S.C. 3685)

1A. TITLE OF PUBLICATION: THE AMERICAN MATHEMATICAL MONTHLY

1B. PUBLICATION NO: A M M Y A E

2. DATE OF FILING: Sept. 30, 1987

3. FREQUENCY OF ISSUE: monthly except bi-monthly June/July and Aug/Sept

3A. NO. OF ISSUES PUBLISHED ANNUALLY: ten

3B. ANNUAL SUBSCRIPTION PRICE: Library $70, Indiv. Member $24

4. COMPLETE MAILING ADDRESS OF KNOWN OFFICE OF PUBLICATION *(Street, City, County, State and ZIP+4 Code) (Not printers)*
1529 Eighteenth Street, N.W., Washington, D.C. 20036

5. COMPLETE MAILING ADDRESS OF THE HEADQUARTERS OF GENERAL BUSINESS OFFICES OF THE PUBLISHER *(Not printer)*
1529 Eighteenth Street, N.W., Washington, D.C. 20036

6. FULL NAMES AND COMPLETE MAILING ADDRESS OF PUBLISHER, EDITOR, AND MANAGING EDITOR *(This item MUST NOT be blank)*

PUBLISHER *(Name and Complete Mailing Address)*
The Mathematical Association of America, 1529 Eighteenth St.,N.W., Washington, D.C. 20036

EDITOR *(Name and Complete Mailing Address)*
Prof. Herbert S. Wilf, U. of Pennsylvania, Dept. of Math., Philadelphia, PA 19104

MANAGING EDITOR *(Name and Complete Mailing Address)*
Harry Waldman, MAA, 1529 Eighteenth St., NW, Washington, D.C. 20036

7. OWNER *(If owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual must be given. If the publication is published by a nonprofit organization, its name and address must be stated.) (Item must be completed.)*

FULL NAME	COMPLETE MAILING ADDRESS
The Mathematical Association of America, Inc.	1529 Eighteenth St., N.W. Washington, D.C. 20036

8. KNOWN BONDHOLDERS, MORTGAGEES, AND OTHER SECURITY HOLDERS OWNING OR HOLDING 1 PERCENT OR MORE OF TOTAL AMOUNT OF BONDS, MORTGAGES OR OTHER SECURITIES *(If there are none, so state)*

FULL NAME	COMPLETE MAILING ADDRESS
None	

9. FOR COMPLETION BY NONPROFIT ORGANIZATIONS AUTHORIZED TO MAIL AT SPECIAL RATES *(Section 423.12 DMM only)*
The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes *(Check one)*

(1) [X] HAS NOT CHANGED DURING PRECEDING 12 MONTHS
(2) [ ] HAS CHANGED DURING PRECEDING 12 MONTHS
(If changed, publisher must submit explanation of change with this statement.)

10. EXTENT AND NATURE OF CIRCULATION *(See instructions on reverse side)*	AVERAGE NO. COPIES EACH ISSUE DURING PRECEDING 12 MONTHS	ACTUAL NO. COPIES OF SINGLE ISSUE PUBLISHED NEAREST TO FILING DATE
A. TOTAL NO. COPIES *(Net Press Run)*	22,732	22,063
B. PAID AND/OR REQUESTED CIRCULATION 1. Sales through dealers and carriers, street vendors and counter sales	100	100
2. Mail Subscription *(Paid and/or requested)*	19,223	18,522
C. TOTAL PAID AND/OR REQUESTED CIRCULATION *(Sum of 10B1 and 10B2)*	19,323	18,622
D. FREE DISTRIBUTION BY MAIL, CARRIER OR OTHER MEANS SAMPLES, COMPLIMENTARY, AND OTHER FREE COPIES	654	958
E. TOTAL DISTRIBUTION *(Sum of C and D)*	19,977	19,580
F. COPIES NOT DISTRIBUTED 1. Office use, left over, unaccounted, spoiled after printing	2,755	2,483
2. Return from News Agents	0	0
G. TOTAL *(Sum of E, F1 and 2—should equal net press run shown in A)*	22,732	22,063

11. I certify that the statements made by me above are correct and complete

SIGNATURE AND TITLE OF EDITOR, PUBLISHER, BUSINESS MANAGER, OR OWNER: Alfred B. Willcox

PS Form 3526, Dec. 1985 *(See instruction on reverse)*

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 95, Number 2 February 1988

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Dept. of Mathematics, Stanford University, Stanford, CA 94305 (address effective until May, 1988). Shorter articles about the teaching of mathematics should be sent to STAN WAGON and JOAN HUTCHINSON, Department of Mathematics, Smith College, Northampton, MA 01063. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to ANITA E. SOLOW, Dept. of Mathematics, Grinnell College, Grinnell, IA 50112; in geometry or topology to DENNIS DETURCK, Dept. of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to RICHARD LIBERA, Mathematical Sciences Department, University of Delaware, Newark, DE 19717. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Proposed problems (three copies) and solutions (two copies), both elementary and advanced to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, University of Calgary, Alberta, Canada T2N 1N4.

Reprint Permission: A. B. WILLCOX, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Back Issues: P. and H. BLISS CO., Middletown, CT 06457.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT

The annual subscription price of $26 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $29. Student, unemployed and emeritus members receive a 50% discount, new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $90 per year

Second class postage paid at Washington, DC, and additional mailing offices

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Assocation of America, 1529 Eighteenth Street, N.W., Washington, DC 20077-9564

PRINTED IN THE UNITED STATES OF AMERICA

All Topologies Come From Generalized Metrics

RALPH KOPPERMAN, *City College of New York*

RALPH KOPPERMAN: I received my Ph.D. from M.I.T. in 1965, and came to The City College in 1967. SEEK is a special program for students from poor areas of the city, and I have coordinated its mathematics efforts since 1969.

My research interest has always been in limits, which I first tried to study through infinitary languages (logic). I was able to publish in the field, but was not happy with the results of that research, and went into point-set topology in 1980. I was a founding member of the CCNY Seminar on General Topology and Topological Algebra in 1981, and have been involved with it ever since.

Most of us learned our topology in a course about metrics. We were shown that the wonderful notions of continuity and uniform continuity, limit and Cauchy sequence, naturally arose from their proper consideration. We were then shown or told that metrics were too restrictive. There were limits which could not arise from metrics.

Topological spaces were given as a partial solution to this problem. More general views of limit and continuity arose from a study of the topological concepts of open and closed, but uniform continuity and the ideas of Cauchy and complete were gone. Depending on our teacher, we may or may not eventually have learned another way of recovering these uniform ideas.

Here we instead generalize the definition of metric space in a way that makes all topologies generalized metric topologies. This approach allows the use of suggestive "metric" notations at all times and reunites the treatment of topological and uniform ideas in key cases.

In the first section, we look at metric and topological spaces from an almost standard viewpoint; in the second we develop our generalization. All topological and metric definitions introduced without comment are as found in Kelley ([**1**]). Similar definitions are found in most basic texts, but there may be some differences.

1. Metric and topological spaces. For our purposes, it is convenient to regard a *metric* as a map $d\colon X \times X \to [0, \infty]$ for some nonempty X, which satisfies:

(m1) $d(x, x) = 0$

(m2) $d(x, z) \leq d(x, y) + d(y, z)$ (triangularity)

(m3) $d(x, y) = d(y, x)$ (symmetry)

(m4) $d(x, y) = 0 \Rightarrow x = y$ (separation).

A *pseudometric* is a map $d\colon X \times X \to [0, \infty]$ for $X \neq \phi$, satisfying (m1)–(m3), a *quasimetric* is one satisfying (m1)–(m2).

A *topology* is a set $\mathbf{T}$ of subsets of some nonempty X such that:
(t1) ϕ, $X \in \mathbf{T}$,
(t2) if $\mathbf{C} \subset \mathbf{T}$ then $\bigcup\mathbf{C} \in \mathbf{T}$,
(t3) if $\mathscr{P}$, $\mathscr{Q} \in \mathbf{T}$ then $\mathscr{P} \cap \mathscr{Q} \in \mathbf{T}$.

Given a topology $\mathbf{T}$, a set is *open* if it's in $\mathbf{T}$, *closed* if its complement is in $\mathbf{T}$.

Just how do we get a metric (or quasimetric) topology? A set $\mathscr{P}$ is open in such a topology if every point in $\mathscr{P}$ is surrounded by an open ball with positive radius r, $S_r(x) = \{y\colon d(x, y) < r\}$ contained in $\mathscr{P}$. It would be equivalent to require that for each $x \in \mathscr{P}$ there is some positive r such that the closed ball, $N_r(x) = \{y\colon d(x, y) \leq r\} \subset \mathscr{P}$: $N_{r/2}(x) \subset S_r(x) \subset N_r(x)$, so if one sort of ball of positive radius is contained in $\mathscr{P}$, so is the other. This latter is easier to generalize, and will be our approach.

We now show that if $\mathscr{X} = (X, d)$ is a metric space, then its metric topology, which we shall denote by $\mathrm{To}(\mathscr{X})$, is a topology:

(t1) ϕ is open since no $X \in \phi$ can contradict our condition. Since there is a positive r, choose one; if $x \in X$ then $N_r(x) \subset X$, showing $X \in \mathrm{To}(\mathscr{X})$.

(t2) If $\mathbf{C} \subset \mathrm{To}(\mathscr{X})$ we show $\bigcup\mathbf{C}$ to be open: if $x \in \bigcup\mathbf{C}$ then for some $\mathscr{Q} \in \mathbf{C}$, $x \in \mathscr{Q}$; but then for some positive r, $N_r(x) \subset \mathscr{Q} \subset \bigcup\mathbf{C}$.

(t3) To show that the intersection of two open sets is open, let $\mathscr{P}, \mathscr{Q} \in \mathrm{To}(\mathscr{X})$, $x \in \mathscr{P} \cap \mathscr{Q}$. Then for some positive r, t, $N_r(x) \subset \mathscr{P}$ and $N_t(x) \subset \mathscr{Q}$. But the minimum of two positives is positive and $N_{\min\{r, t\}}(x) = N_r(x) \cap N_t(x) \subset \mathscr{P} \cap \mathscr{Q}$, so we're done.

The two concepts of topology and metric give us two notions of limit: Let (X, d) and (Y, d') be metric spaces. L is a *metric limit* at $a \in X$ for a function $f\colon X \to Y$ if for each positive r there is a positive s such that if $d(a, x) \leq s$ then $d'(L, f(x)) \leq r$.

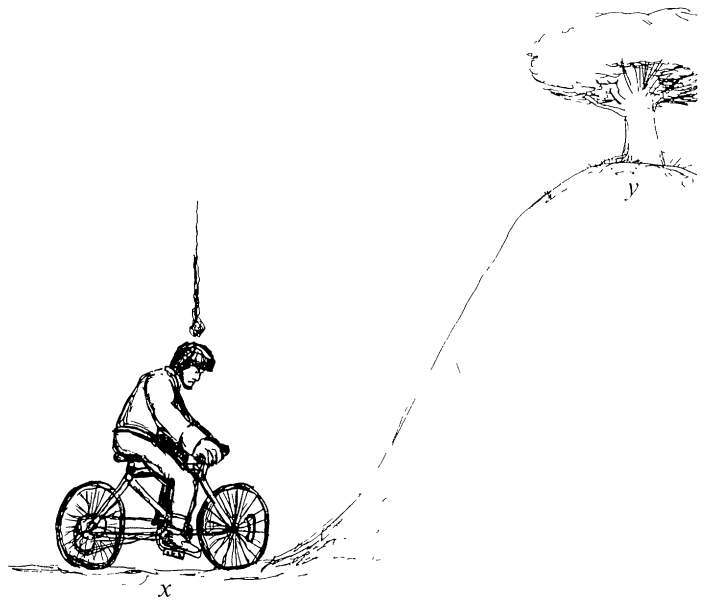
For $(X, \mathbf{T})$, $(Y, \mathbf{T}')$ topological spaces, L is a *topological limit* at $a \in X$ for a function $f\colon X \to Y$ if for each open $\mathscr{Q}$, if $L \in \mathscr{Q}$ there is an open $\mathscr{P}$ such that $a \in \mathscr{P}$ and for each $x \in \mathscr{P}$, $f(x) \in \mathscr{Q}$.

We leave to the reader the important verification that if $\mathbf{T}$ is the topology arising from d and $\mathbf{T}'$ arises from d', then metric limits and topological limits are the same.

What must we change? Nonmathematicians would accept the cartoon (p. 91) as a demonstration of nonsymmetry.

But mathematicians require more, so we look at two topologies indirectly discussed in some calculus courses.

The topology of upper semicontinuity. Recall that $f\colon \mathbf{R} \to \mathbf{R}$ ($\mathbf{R}$ the reals) is *upper semicontinuous* if whenever $f(x) < a$ then for some positive r, if $|x - y| < r$ then $f(y) < a$. It can be shown that the collection $\mathbf{U} = \{(-\infty, a)\colon -\infty \leq a \leq \infty\}$ forms a topology on $\mathbf{R}$, called *the topology of upper semicontinuity*, and that f is upper semicontinuous iff f is continuous from the usual real topology to $\mathbf{U}$ (for discussion see Kelley [1], p. 101). In $\mathbf{U}$ if $x < y$ and $\mathscr{P} = (-\infty, y)$ then $x \in \mathscr{P}$ and $y \notin \mathscr{P}$ but there is no open $\mathscr{Q}$ such that $y \in \mathscr{Q}$ and $x \notin \mathscr{Q}$ (for if $y \in \mathscr{Q} = (-\infty, z)$



The cynical cyclist and symmetry.

then $x < y < z$ so $x \in \mathscr{Q}$). But:

Fact. In the topology arising from a quasimetric d on X there is an open $\mathscr{P}$ with $x \in \mathscr{P}$ and $y \notin \mathscr{P}$ iff $0 < d(x, y)$.

Proof. If $d(x, y) = 0$ and $x \in \mathscr{P}$, $\mathscr{P}$ open, let $N_r(x) \subset \mathscr{P}$; then $d(x, y) \leq r$ so $y \in \mathscr{P}$. But if $d(x, y) > 0$ then let $r > 0$, $d(x, y) \nleq r$; clearly $x \in S_r(x)$, $y \notin S_r(x)$, an open set. (The usual proof shows that $S_r(x)$ is open in this situation; a generalization of that result is shown in the next section.)

Combining this fact with the previous discussion of $\mathbf{U}$, we see that if $\mathbf{U}$ arises from a quasimetric d and $x < y$, then $d(y, x) = 0$ contradicting (m4), and $d(x, y) > 0 = d(y, x)$, contradicting (m3) as well. (The reader may wish to verify a fact not central to our discussion: $\mathbf{U}$ arises from the quasimetric defined by $d(x, y) = \max\{0, y - x\}$.)

The topology of pointwise convergence and neighborhood bases. Given a topology $\mathbf{T}$, and an $x \in X$, a *neighborhood base about* x is a collection $\mathbf{B}_x$ of sets which have open subsets containing x and for which if $x \in \mathscr{P}$, $\mathscr{P}$ open, then for some $N \in \mathbf{B}_x$, $N \subset \mathscr{P}$. Note that for any metric topology, if $x \in X$, $\mathbf{B}_x = \{S_{1/n}(x)$: n a positive integer$\}$ is a countable base about x. Further, this collection is totally ordered by set inclusion since if $r \leq s$ then $S_r(x) \subset S_s(x)$.

Let $X = \mathbf{R}^{[0,1]} = \{f\colon [0,1] \to \mathbf{R}\}$, under the topology of pointwise convergence. (Recall that a sequence $\{g_n\}$ of functions from Y into $\mathbf{R}$ *converges pointwise* to a

function f if $\{g_n(y)\}$ converges to $f(y)$ for each $y \in Y$.) Kelley ([**1**], 92–93 particularly Theorem 6, also 88–91) discusses the topology of pointwise convergence and shows that points in this X don't have countable bases about them.

For us it's useful to note further that points don't even have bases about them totally ordered by set inclusion. Let $f \in \mathbf{R}^{[0,1]}$ and define $f_n \in \mathbf{R}^{[0,1]}$ by $f_n(x) = f(x) + 1/n$; then $\{f_n\}$ is a sequence approaching f pointwise. By way of contradiction let $\mathbf{B}$ be a base about f totally ordered by set inclusion, and for each n let $\mathscr{P}_n \in \mathbf{B}$, $f_n \notin \mathscr{P}_n$. If there were a $\mathscr{Q} \in \mathbf{B}$ such that for each n $\mathscr{Q} \subset \mathscr{P}_n$, then for each n, $f_n \notin \mathscr{Q}$, and since $\mathscr{Q}$ is a neighborhood of f, this contradicts $\{f_n\} \to f$. Thus for some n, $\mathscr{Q} \not\subset \mathscr{P}_n$, and since $\mathbf{B}$ is totally ordered by set inclusion, $\mathscr{P}_n \subset \mathscr{Q}$. But then if $f \in \mathscr{S}$, $\mathscr{S}$ open, there is a $\mathscr{Q} \in \mathbf{B}$ such that $\mathscr{Q} \subset \mathscr{S}$ and an n such that $\mathscr{P}_n \subset \mathscr{Q}$, thus $\mathscr{P}_n \subset \mathscr{S}$, so $\{\mathscr{P}_n\colon n = 1, 2, \dots\}$ is a countable base about f, our contradiction. Thus this space cannot arise from a metric valued in any totally ordered set like the nonnegative reals. We've now seen the key reasons why not all topologies arise from metrics, and shown how an idea of limit comes from topology.

We close this section with an example showing that a topology doesn't determine its Cauchy sequences: Let X be the set of positive integers, $d(m, n) = |m - n|$, $e(m, n) = |1/m - 1/n|$ $(= d(m, n)/mn)$. Both d and e are metrics on X, and further $\mathrm{To}((X, d)) = \mathrm{To}((X, e))$: since $S_1^d(n) = S_{1/n(n+1)}^e(n) = \{n\}$, $\{n\}$ is open in each topology, so since arbitrary unions of open sets are open and each set is the union of its points, each set is open in either topology.

Thus if the notion of Cauchy were determined by topology, the two metrics should yield the same Cauchy sequences. Since $d(m, n) \geq 1$ for $m \neq n$, $x_n = n$ isn't Cauchy with respect to d. But it is with respect to e.

2. Continuity spaces. In the first section we showed that (m3) and (m4) (symmetry and separation) couldn't be expected to hold in a generalization of metric space from which all topologies arose. Neither could we expect our generalized metric to be valued in the positive reals. We begin this section with a definition of structures which may be used to replace the nonnegative extended reals.

Before we give our definition, notice that the order on $[0, \infty]$ is unusual in that it's completely determined by addition: $a \leq b$ iff for some c $(\in [0, \infty])$, $b = a + c$. This in turn says about $+$ on $[0, \infty]$ that if $a + x = b$ and $b + y = a$ then $a \leq b$ and $b \leq a$ so $a = b$. We adopt this as our axiom (v1) below. Notice that if $(A, +)$ is a semigroup (i.e., $+\colon A \times A \to A$ is associative) and A has identity 0 then $\leq$ can be defined on A by: $a \leq b$ iff for some $x \in A$, $b = a + x$. This $\leq$ is always reflexive (since $a = a + 0$, $a \leq a$) and transitive (if $a \leq b$ and $b \leq c$ then for some $x, y \in A$, $b = a + x$ and $c = b + y = (a + x) + y = a + (x + y)$, so $a \leq c$). If (v1) holds, $\leq$ becomes a partial order (since if $a \leq b$ and $b \leq a$ then for some $x, y \in A$, $b = a + x$ and $a = b + y$ so $a = b$). Later we give examples which show that this $\leq$ need not be total.

In giving up our total order, we must give up the existence of $\min\{r, s\}$ for $r, s > 0$, a number often used in metric arguments; we shall find that inf will work

in its place. With this preface, we now generalize $[0, \infty]$:

1. *Definition*. A *value semigroup* is an additive abelian semigroup A with identity 0 and absorbing element $\infty \neq 0$, satisfying:
(v1) if $a + x = b$ and $b + y = a$ then $a = b$,
(thus $a \leq b$ iff $b = a + x$ for some x, defines a partial order),
(v2) for each a there is a unique b (called $(1/2)a$ or $a/2$) such that $b + b = a$,
(v3) for each a, b there is an inf, $a \wedge b$,
(v4) $a \wedge b + c = (a + c) \wedge (b + c)$.

Fact. If for each $i \in I$, A_i is a value semigroup (together with $+_i$, 0_i, ∞_i), then so is their product (with $+$, 0, defined coordinatewise; 1/2 and inf are also taken coordinatewise).

Proof. (v1) Suppose $a + x = b$, $b + y = a$. Then for each i, $a_i +_i x_i = (a + x)_i = b_i$, $a_i = b_i +_i y_i$, so $a_i = b_i$, so $a = b$.
(v2)–(v4) are shown similarly and their proofs are left to the reader.

The nonnegative extended reals with the usual addition form an example of a value semigroup which we call $\mathscr{R}$. The examples we use later will be powers of $\mathscr{R}$.
The partial order $\leq$ defined using (v1) is the usual one on $[0, \infty]$, and is the only one we refer to below. For each $a, b \in A$, $0 \leq a/2 \leq a \leq a + b \leq \infty$. Thus thinking of $<$ as $\leq$ and $\neq$, each nonzero element is positive.
In $\mathscr{R}^2$, $(0,1) \not\leq (1,0)$ and $(1,0) \not\leq (0,1)$; in fact $(0,1)$ and $(1,0)$ are two positive elements whose inf is 0. Many metric space proofs use the fact that the min of two positive numbers is positive, in fact this property of the positive reals, rather than their size relative to 0 is key to their usefulness. Thus we must next define a "set of positives" by its useful properties:

2. *Definition*. A *set of positives* in a value semigroup A is a $P \subset A$ satisfying:
(p1) if $r, s \in P$ then $r \wedge s \in P$,
(p2) if $r \in P$, $r \leq a$, then $a \in P$,
(p3) if $r \in P$ then $r/2 \in P$,
(p4) if $a \leq b + r$ for each $r \in P$ then $a \leq b$.

(p4) is often used in its contrapositive form: if $a \not\leq b$ then $a \not\leq b + r$ for some $r \in P$. Also $P \neq \phi$ by (p4), for otherwise $\infty \leq 0 + r$ for each $r \in P$ so $\infty \leq 0$, and since $\leq$ is a partial order and $0 \leq \infty$ by the paragraph preceding definition 2, $0 = \infty$; but this contradicts definition 1. In addition to the usual example, $P = (0, \infty]$ in $\mathscr{R}$, we see that each value semigroup A is a set of positives in itself. We can now define our generalization of metric space:

3. *Definition*. A *continuity space* is a quadruple $\mathscr{X} = (X, d, A, P)$ such that X is nonempty, A is a value semigroup, P a set of positives on A, and d: $X \times X \to A$ satisfies (m1) and (m2) of our first paragraph. This d is called a *continuity function*. $\mathscr{X}$ (and d) is *symmetric* iff (m3) is satisfied, and *separated* iff (m4) is satisfied.
We now formalize the way topological spaces arise from continuity spaces.

4. *Definition*. If $x \in X$, $b \in A$, $\mathscr{X} = (X, d, A, P)$ a continuity space, then $N_b(x) = \{y : d(x, y) \le b\}$ is called the *closed ball* of radius b about x, and $\mathrm{To}(\mathscr{X}) = \{\mathscr{Q} \subset X$: if $x \in \mathscr{Q}$ then $N_r(x) \subset \mathscr{Q}$ for some $r \in P\}$ is the *topology arising from* $\mathscr{X}$.

Suggestive "metric" notation is valid for continuity space topologies. For example, if f: $X \to Y$, $\mathscr{X} = (X, d, A, P)$, $\mathscr{Y} = (Y, e, A', P')$, then f is continuous at $x \in X$ with respect to $\mathrm{To}(\mathscr{X})$, $\mathrm{To}(\mathscr{Y})$ iff for each $s \in P'$ there is an $r \in P$ such that if $d(x, y) \le r$ then $e(f(x), f(y)) \le s$. Uniform continuity can be defined in the above situation as well: f is *uniformly continuous* iff for each $s \in P'$ there is an $r \in P$ such that if $d(x, y) \le r$ then $e(f(x), f(y)) \le s$ (here r is independent of x). In order to discuss certain limits in this general setting, we need to generalize sequences (mappings from the natural numbers into X) to *nets* (maps from a directed set $(D, \le)$ into X; Kelley discusses these beginning on p. 65). Then y is the limit of the net x: $D \to X$ iff for each $r \in P$ there is an $n \in D$ such that if $n \le m$ then $d(y, x_m) \le r$. We can also define the net x: $D \to X$ to be *Cauchy* if for each $r \in P$ there is an $n \in D$ such that if $n \le m, p$ then $d(x_m, x_p) \le r$. It's a useful exercise to generalize the usual textbook results about topological concepts (limit, continuous, open, closed) and uniform notions (uniformly continuous, Cauchy, complete) to this setting. Most topological results hold without additional restrictions; some of them, as well as most uniform results, require symmetry, and separation is sometimes a simplifying assumption. (Note: there is also a clear way to define a uniformity from a continuity space in the symmetric case. If this is done, our definitions of uniform concepts coincide with those which come from the associated uniformity, but we prefer not to discuss uniform spaces here. The place of symmetry in our scheme is given in Theorem 11.) The Baire category theorem and related results are the few which require that we essentially have a metric (valued in $[0, \infty)$).

5. PROPOSITION. $\mathrm{To}(\mathscr{X})$ *is always a topology.*

Proof. Our proof in the first section that every metric yields a topology applies here word for word, except for its last sentence. Note that since $d(x, y) \le r \wedge t$ iff $d(x, y) \le r$ and $d(x, y) \le t$, $N_{r \wedge t}(x) = N_r(x) \cap N_t(x)$. Induction extends this to finite infs and intersections. That last sentence may now be changed to read: But since the inf of two positives is positive (by (p1)), $N_{r \wedge t}(x) = N_r(x) \cap N_t(x) \subset \mathscr{P} \cap \mathscr{Q}$, so we've shown $\mathrm{To}(\mathscr{X})$ to be a topology.

Our main goal is to show the converse of 5: every topology is a $\mathrm{To}(\mathscr{X})$. It's useful first to look at what needs to be done. For $(X, \mathbf{T})$ a topological space, we must find A, P, d so that if $\mathscr{X} = (X, d, A, P)$ then $\mathbf{T} = \mathrm{To}(\mathscr{X})$. But how do we know when $\mathbf{T} = \mathrm{To}(\mathscr{X})$? Clearly by definition 4, if for each $x \in \mathscr{O}$ there's an $r \in P$ such that $N_r(x) \subset \mathscr{O}$, then $\mathscr{O} \in \mathrm{To}(\mathscr{X})$, so if this holds for each $\mathscr{O} \in \mathbf{T}$ we must have $\mathbf{T} \subset \mathrm{To}(\mathscr{X})$.

It isn't difficult to get for $\mathscr{S} \in \mathrm{To}(\mathscr{X})$ a continuity function $d_{\mathscr{S}}$: $X \times X \to \mathscr{R}$ such that $\mathscr{S} \in \mathrm{To}(\mathscr{X}_{\mathscr{S}}) \subset \mathrm{To}(\mathscr{X})$, where $\mathscr{X}_{\mathscr{S}} = (X, d_{\mathscr{S}}, \mathscr{R}, (0, \infty])$. Simply let $q > 0$ and define $d_{\mathscr{S}}(x, y) = 0$ if $(x \in \mathscr{S} \Rightarrow y \in \mathscr{S})$, $d_{\mathscr{S}}(x, y) = q$ otherwise. (m1) is

clearly satisfied, and for (m2): If $d_{\mathscr{S}}(x, y) = q$ or $d_{\mathscr{S}}(y, z) = q$ then $d_{\mathscr{S}}(x, z) \leq d_{\mathscr{S}}(x, y) + d_{\mathscr{S}}(y, z)$, and if both are 0 then $(x \in \mathscr{S} \Rightarrow y \in \mathscr{S})$ and $(y \in \mathscr{S} \Rightarrow z \in \mathscr{S})$ so $(x \in \mathscr{S} \Rightarrow z \in \mathscr{S})$, showing $d_{\mathscr{S}}(x, z) = 0 \leq d_{\mathscr{S}}(x, y) + d_{\mathscr{S}}(y, z)$. Also if $p < q$ then $N_p(x) = \{y\colon d_{\mathscr{S}}(x, y) \leq p\} = \{y\colon d_{\mathscr{S}}(x, y) = 0\} = \{y\colon x \in \mathscr{S} \Rightarrow y \in \mathscr{S}\}$; this set is $\mathscr{S}$ if $x \in \mathscr{S}$, X otherwise, and clearly if $p \geq q$ then $N_r(x) = X$. By 5, ϕ, $X \in \mathrm{To}(\mathscr{X}_{\mathscr{S}})$; if $x \in \mathscr{S}$, $0 < p < q$ then $N_p(x) \subset \mathscr{S}$, thus $\mathscr{S} \in \mathrm{To}(\mathscr{X}_{\mathscr{S}})$. If $\mathscr{P} \in \mathrm{To}(\mathscr{X}_{\mathscr{S}})$ and $\mathscr{P} \neq \phi$ then let $x \in \mathscr{P}$; there's some $p > 0$ such that $\mathscr{S} \subset N_p(x) \subset \mathscr{P}$, thus $\mathscr{S} \subset \mathscr{P}$. If $\mathscr{S} \neq \mathscr{P}$ we can choose $x \in \mathscr{P} - \mathscr{S}$ and then $X = N_p(x) \subset \mathscr{P}$, so $X = \mathscr{P}$. Thus $\mathrm{To}(\mathscr{X}_{\mathscr{S}}) = \{0, \mathscr{S}, X\} \subset \mathrm{To}(\mathscr{X})$.

The above suggests that we might well try to build up from an appropriate collection of $\mathscr{X}_{\mathscr{S}}$'s, and the next lemma sets up equipment which will allow us to do that.

6. Lemma. *If for each $i \in I$, A_i is a value semigroup and $P_i \subset A_i$ is a set of positives, then the* sum *of the P_i, $P = \{r\colon r(i) \in P_i$ for each $i \in I$ and for all but a finite number of $i \in I$, $r(i) = \infty_i\}$ is a set of positives in the product A of the A_i's.*

Proof. (p1) if $r, s \in P$, then for each $i \in I$, $(r \wedge s)(i) = r(i) \wedge s(i) \in P_i$, and $= \infty_i$ for all but the finite number of places at which $r(i) \neq \infty_i$ or $s(i) \neq \infty_i$, thus $r \wedge s \in P$; (p2) and (p3) are clear. For (p4) assume $b \leq a + r$ for each $r \in P$, and note that if $i \in I$, $q \in P_i$ and $r(i, q) \in P$ is defined by $r(i, q)(i') = \infty_i$ if $i \neq i' \in I$, $r(i, q)(i) = q$, then $r(i, q) \in P$. Thus $b(i) \leq (a + r)(i) = a(i) + q$ for arbitrary positive q, so $b(i) \leq a(i)$, but since i was also arbitrary, $b \leq a$.

7. Theorem. *Every topology is a* $\mathrm{To}(\mathscr{X})$.

Proof. Let $q \in (0, \infty]$ be fixed throughout this proof. For each $\mathscr{S} \in \mathbf{T}$ (our topology) define $d_{\mathscr{S}}$ as preceding 6 and let $\mathscr{X} = (X, d, A, P)$, where A is the product of $\mathbf{T}$ copies of $\mathscr{R}$, P the sum of $\mathbf{T}$ copies of $(0, \infty]$, and for each $x, y \in X$, $\mathscr{S} \in \mathbf{T}$, $d(x, y)(\mathscr{S}) = d_{\mathscr{S}}(x, y)$. A was shown to be a value semigroup following definition 1; P was shown to be a set of positives on A in proposition 6. Further, we showed that (m1), (m2) were satisfied at each coordinate by $d_{\mathscr{S}}$, thus they are satisfied by $\mathscr{X}$. It remains only to be shown that for $\mathscr{X}$ defined as above, $\mathrm{To}(\mathscr{X}) = \mathbf{T}$.

Let $\mathscr{S} \in \mathbf{T}$, $p \in (0, \infty]$ and define $r(\mathscr{S}, p)$ by $r(\mathscr{S}, p)(\mathscr{S}) = p$, $r(\mathscr{S}, p)(\mathscr{S}') = \infty$ if $\mathscr{S} \neq \mathscr{S}'$. $N_{r(\mathscr{S}, p)}(x) = \{y\colon d(x, y) \leq r(\mathscr{S}, p)\} = \{y\colon d_{\mathscr{S}}(x, y) \leq p\}$, thus by the discussion of $d_{\mathscr{S}}$, this set is $\mathscr{S}$ if $x \in \mathscr{S}$ and $p \leq q$, X otherwise. This shows that $\mathbf{T} \subset \mathrm{To}(\mathscr{X})$, since if $x \in \mathscr{S}$, then for $p < q$, $r(\mathscr{S}, p) \in P$ and $N_{r(\mathscr{S}, p)}(x) \subset \mathscr{S}$.

It further leads to $\mathrm{To}(\mathscr{X}) \subset \mathbf{T}$: If $x \in \mathscr{P} \in \mathrm{To}(\mathscr{X})$ then for some $r_x \in P$, $N_{r_x}(x) \subset \mathscr{P}$ but $\{\mathscr{S}\colon r_x(\mathscr{S}) < \infty\}$ is finite, thus let $r_x(\mathscr{S}_i) = p_i$ for $1 \leq i \leq n$, $r_x(\mathscr{S}') = \infty$ for $\mathscr{S}' \notin \{\mathscr{S}_1, \ldots, \mathscr{S}_n\}$. Then $r_x = r(\mathscr{S}_1, p_1) \wedge \cdots \wedge r(\mathscr{S}_n, p_n)$, thus $N_{r_x}(x)$ is the finite intersection of the $N_{r(\mathscr{S}_i, p_i)}(x)$. Since each of these is X or $\mathscr{S}_i$, $N_{r_x}(x) \in \mathbf{T}$. Thus $\mathscr{P} \subset \bigcup_{x \in \mathscr{P}} N_{r_x}(x) \subset \mathscr{P}$, so $\mathscr{P}$ is a union of elements of $\mathbf{T}$, therefore by (t2), $\mathscr{P} \in T$.

3. Symmetry and complete regularity. The above raises obvious questions as to the place of symmetry and (m4) in our scheme. A surprising feature of the proof of 7 is that our "closed balls" were open, but not closed, and we also discuss questions related to that. Recall that $\mathbf{T}$ is T_1 iff for each x, $\{x\}$ is closed ($X - \{x\}$ is open),

completely regular iff for each open $\mathscr{P}$ and each $x \in \mathscr{P}$ there is a continuous $f\colon X \to [0,1]$ such that $f(x) = 1$ and $f(y) = 0$ for each $y \notin \mathscr{P}$, *Tychonoff* iff it is T_1 and completely regular.

8. PROPOSITION. $\mathrm{To}(\mathscr{X})$ *is* T_1 iff $\mathscr{X}$ *is separated.*

Proof. If $\mathscr{X}$ is separated and $x \in X - \{y\}$ then $d(x, y) \neq 0$ so for some $r \in P$, $d(x, y) \nleq r$; for this r, $N_r(x) \subset X - \{y\}$, and since x was arbitrary, $X - \{y\}$ is open. Conversely, if $\mathscr{X}$ is T_1 and $x, y \in X$, $x \neq y$, then $x \in X - \{y\}$ so for some $r \in P$, $N_r(x) \subset X - \{y\}$, thus $d(x, y) \nleq r$ and in particular, $d(x, y) \neq 0$.

Next, open and closed balls. Note that the obvious generalization of $S_r(x)$ need not always be an open set: Let X be the plane, $A = \mathscr{R}^2$, $P = (0, \infty]^2$, and $d((w, x), (y, z)) = (|w - y|, |x - z|)$, $\mathscr{X} = (X, d, A, P)$. The reader can verify that $\mathscr{X}$ is a continuity space and $\mathrm{To}(\mathscr{X})$ is the usual topology, but $S_{(1,1)}((0, 0))$ is a closed square with the corners removed, thus not open. Some other concepts are needed:

9. *Definition*. The *positive dyadics* are $DY = \{n = i/2^k\colon i, k \text{ positive integers}\}$.

Notice that there is a multiplication $\cdot : DY \times A \to A$ defined by repeated application of addition and $1/2$ (of (v2)).

For $x \in X$, $b \in A$, $B_b(x) = \bigcup\{N_{nb}(x)\colon n \in DY,\ n < 1\} = \{y\colon \text{for some dyadic } n < 1,\ d(x, y) \leq nb\}$, and is called the *open ball of radius b about x*.

The *dual* of $\mathscr{X} = (X, d, A, P)$ is $\mathscr{X}^* = (X, d^*, A, P)$, where $d^*(x, y) = d(y, x)$ (and X, A, P remain unchanged). Dual concepts are denoted by placing $*$ in appropriate places, e.g., $N_r^*(x) = \{y\colon d^*(x, y) \leq r\}$.

It can be shown routinely that this dyadic multiplication is well defined, distributive in both variables, $(mn)a = m(na)$ and $1a = a$. Thus if $n < m$ then $nr \leq nr + (m - n)r = (n + m - n)r = mr$. It's also clear by (p3) and (p2) that if $r \in P$ and $n \in DY$ then $nr \in P$. Further, d^* is always a continuity function, thus $\mathscr{X}^*$ a continuity space, and for metric spaces ($A = \mathscr{R}$, $P = (0, \infty]$), if $r \in P$ then $S_r(x) = B_r(x)$.

10. PROPOSITION. *If* $b \in A$, $x \in X$ *then* $N_{b/2}(x) \subset B_b(x) \subset N_b(x)$. $B_r(x)$ *is open if* $r \in P$. *For each* $x \in X$, *the sets* $\mathbf{B}_x = \{B_r(x)\colon r \in P\}$ *and* $\mathbf{N}_x = \{N_r(x)\colon r \in P\}$ *are both bases about x. For each* $b \in A$, $x \in X$, $N_b^*(x)$ *is closed in* $\mathrm{To}(\mathscr{X})$.

Proof. The first inclusion holds since $b/2 = (1/2)b$, and the second follows from the fact that if $n < 1$ then $nb \leq 1b = b$, so $N_{nb}(x) \subset N_b(x)$. We next show that if $r \in P$ then $B_r(x)$ is open: if $y \in B_r(x)$ then for some $n < 1$, $d(x, y) \leq nr$; if $m = (1 - n)/2$ then $m \in DY$ so $mr \in P$, and if $z \in N_{mr}(y)$ then $d(x, z) \leq d(x, y) + d(y, z) \leq (n + m)r$, so $z \in N_{(n+m)r}(x)$, and $n + m < 1$, showing $N_{mr}(y) \subset B_r(x)$. That $\mathbf{B}_x$ and $\mathbf{N}_x$ are bases about x now follows, since if $x \in \mathscr{P} \in \mathrm{To}(\mathscr{X})$ then for some $r \in P$, $x \in B_r(x) \subset N_r(x) \subset \mathscr{P}$.

Finally, to see that $N_r^*(x)$ is closed, suppose that $y \in X - N_r^*(x)$. Then $d(y, x) \nleq r$ so by (p4), for some $t \in P$, $d(y, x) \nleq r + t$. If $d(y, z) \leq t$ then $d(z, x) \nleq r$, for otherwise $d(y, x) \leq d(y, z) + d(z, x) \leq t + r$, a contradiction. But this contradiction shows $N_t(y) \subset X - N_r^*(x)$, thus the latter is open.

11. THEOREM. *If $\mathscr{X}$ is symmetric then* $\mathrm{To}(\mathscr{X})$ *is a completely regular topology, which is Tychonoff* iff $\mathscr{X}$ *is separated. Every completely regular topology arises from a symmetric continuity space.*

Proof. Suppose first that $\mathscr{X}$ is symmetric and let $x \in \mathscr{P} \in \mathrm{To}(\mathscr{X})$. Then for some $r \in P$, $N_r(x) \subset \mathscr{P}$. Define $g_r\colon A \to [0, \infty]$ by $g_r(a) = \inf\{n \in DY\colon a \leq nr\}$; note that if $c \leq a + b$ and $a \leq nr$, $b \leq mr$ then $c \leq nr + mr = (n + m)r$, so $g_r(c) \leq g_r(a) + g_r(b)$. We define $h\colon X \to [0,1]$ by $h(y) = \min\{g_r(d(x, y)), 1\}$ and next show that h is continuous.

Since $d(x, z) \leq d(x, y) + d(y, z)$ we have $h(z) \leq h(y) + g_r(d(y, z))$ so $h(z) - h(y) \leq g_r(d(y, z))$; similarly, $h(y) - h(z) \leq g_r(d(z, y))$, so by symmetry, $|h(y) - h(z)| \leq g_r(d(y, z))$. If $p \in (0, \infty]$ then let $n \in DY$, $n < p$; if $d(y, z) \leq nr$ then $|h(y) - h(z)| \leq g_r(d(x, y)) \leq n < p$, so h is continuous at this arbitrary y, thus on X.

Also, $h(x) = \inf DY = 0$ and if $h(y) < 1$ then for some $n < 1$, $d(x, y) \leq nr \leq nr + (1 - n)r = (n + 1 - n)r = r$, so $y \in N_r(x) \subset \mathscr{P}$. Thus $f = \max\{0, 1 - h\}\colon X \to [0,1]$ is our function.

We now proceed to show the converse; note in passing similarities between this proof and that of Theorem 7. Suppose $\mathbf{T}$ is completely regular; then there is a set $\mathscr{F}$ of continuous functions valued in $[0,1]$ such that whenever $x \in \mathscr{P} \in \mathrm{To}(\mathscr{X})$ there is an $f \in \mathscr{F}$ for which $f(x) = 1$ and f is 0 off $\mathscr{P}$. We define A and P and show them to have the right properties in a manner analogous to that used in the proof of 7: $A = \mathscr{R}^{\mathscr{F}}$, $P = \{r \in (0, \infty]^{\mathscr{F}}\colon r(f) = \infty$ for all but a finite number of $f\}$. The definition of d must again be coordinatewise: $d(x, y)(f) = |f(x) - f(y)|$, and there is an obvious coordinatewise proof that d satisfies (m1)–(m3). We now show that $\mathbf{T} = \mathrm{To}(\mathscr{X})$.

For $r \in P$ let $r(f_i) = p_i > 0$ for $1 \leq i \leq n$, $r(g) = \infty$ otherwise; then $y \in B_r(x) \Leftrightarrow d(x, y) \leq mr$ for some $m < 1$, $m \in DY \Leftrightarrow$ for each i between 1 and n, $|f_i(y) - f_i(x)| \leq mp_i$ for some $m \in DY$, $m < 1 \Leftrightarrow$ for each i between 1 and n, $|f_i(y) - f_i(x)| < p_i \Leftrightarrow y \in \bigcap_{i=1}^{n} f_i^{-1}[(f_i(x) - p_i, f_i(x) + p_i)]$. This shows that

$$\left\{ \bigcap_{i=1}^{n} f_i^{-1}[(f_i(x) - p_i, f_i(x) + p_i)]\colon x \in X,\ f_1, \dots, f_n \in \mathscr{F}, \qquad p_1, \dots, p_n > 0\right.$$

n a positive integer$\} = \{B_r(x)\colon x \in X\}$, so for each $r \in P$, $x \in X$, $B_r(x) \in \mathbf{T}$. If $x \in \mathscr{P} \in \mathrm{To}(\mathscr{X})$ choose $r(x) \in P$ such that $B_{r(x)}(x) \subset \mathscr{P}$; then $\mathscr{P} \subset \bigcup_{x \in \mathscr{P}} B_{r(x)}(x) \subset \mathscr{P}$, so $\mathscr{P} = \bigcup_{x \in \mathscr{P}} B_{r(x)}(x) \in \mathbf{T}$ as a union of sets in $\mathbf{T}$, so $\mathrm{To}(\mathscr{X}) \subset \mathbf{T}$. But if $x \in \mathscr{P} \in \mathbf{T}$ choose $f \in \mathscr{F}$ such that $f(x) = 1$, $f(y) = 0$ for $y \notin \mathscr{P}$ and $r(f) \in P$ by $r(f)(f) = .5$, $r(f)(g) =$ for $g \neq f$, and note that $N_{r(f)}(x) = \{y\colon d(x, y) \leq r(f)\} = \{y\colon |1 - f(y)| \leq .5\} \subset \mathscr{P}$, so $\mathscr{P} \in \mathrm{To}(\mathscr{X})$, so $\mathrm{To}(\mathscr{X}) = \mathbf{T}$.

The assertions about Tychonoff spaces are verified by combining the above with Proposition 8.

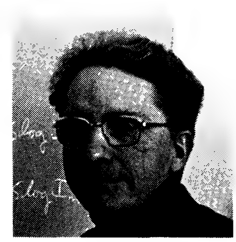
REFERENCE

1. J. L. Kelley, General Topology, Van Nostrand, New York, 1955.

Prime Desert *n*-Tuplets

George W. Polites, *Wesleyan University*

George W. Polites did his graduate work at Florida State University and the University of Illinois. His interests include history of mathematics, and he recently took a group of students to Greece to study the history of Greek mathematics.



Although the number of primes is infinite, one can still find as many consecutive composite integers "as one pleases"; 100, 500, 1,000, 1,000,000, or whatever. That is, there are "regions" of consecutive integers in the set $N = \{1, 2, 3, \dots\}$ of positive integers where no prime is to be found and the number of integers in such a region can be as many as we please.

For example, if we wish to point out a region of 1,000 consecutive composite integers in N we may consider the numbers $1001! + 2, 1001! + 3, \dots, 1001! + 1001$ and observe that the first integer is divisible by 2, the second is divisible by 3, and so on, and that the last integer is divisible by 1001. In general, to find k consecutive positive integers in N, none of which is a prime, we need only consider the numbers $(k+1)! + 2,\ (k+1)! + 3, \dots, (k+1)! + (k+1)$. It is to be pointed out, of course, that these numbers may not form the *first* such region in N. For example, the numbers $22! + 2, 22! + 3, \dots, 22! + 22$ form a region of 21 consecutive integers where no prime is present, but the integers $1130, 1131, \dots, 1150$ form the first such region. It is also to be pointed out that one or both of the numbers $(k+1)! + 1$, $(k+1)! + (k+2)$ may not be prime, meaning that the region can be extended in one direction or the other (perhaps both) to include additional consecutive composite integers. This is not the case for the region containing the integers from 1130 to 1150, inclusive, since 1129 and 1151 are both prime numbers.

When a region of k consecutive composite integers $c_1, c_2, \dots, c_k$ cannot be extended in either direction to include additional composites because the integers $c_1 - 1$ and $c_k + 1$ are both primes, we shall call the region a *prime desert of length k*. Thus the subset $\{1130 + 1, 1131, \dots, 1149, 1150\}$ of N is a prime desert of length 21 (and it is the first prime desert of length 21).

Several questions immediately arise:

1. Is there a prime desert of any given length k?
2. If there is a prime desert of length k, how can one be found? Can we "construct" one?

3. If there is a prime desert of length k, where is the first such desert?
4. If there is a prime desert of length k, is the number of such deserts finite or infinite?

With regard to question 1 we can see immediately that in order to have a prime desert of length k, k must be an *odd* integer. So question 1 becomes: Is there a prime desert of any given odd length k?

With regard to question 2 we remark again that the numbers $(k + 1)! + 2$, $(k + 1)! + 3, \ldots, (k + 1)! + (k + 1)$ are k consecutive composite integers. But they do *not* form a prime desert of length k because one of the integers $(k + 1)! + 1$ or $(k + 1)! + (k + 2)$ must be a composite, as we now show. We observe first that in order for $(k + 1)! + (k + 2)$ to be prime, k will have to be an odd integer such that $k + 2$ is a prime. For suppose k is odd such that $k + 2$ is *not* a prime. Then $k + 2$ has a prime factor q, where $q \leqslant \sqrt{k+2}$. Since $\sqrt{k+2} < k + 1, q < k + 1$ and it follows that $q|(k + 1)!$. Thus $(k + 1)! + (k + 2)$ is not a prime. Similarly, if $k + 4$ is not a prime, then $(k + 1)! + (k + 4)$ is not a prime. In general, we have the following:

THEOREM 1. *Suppose k is odd and consider the set $S = \{k + 2, k + 4, \ldots, k + 2m\}$ where it is the case that each integer in S is not a prime but $k + 2(m + 1)$ is a prime. If $k^2 + k > 2m - 1$, then the consecutive composites $(k + 1)! + 2, (k + 1)! + 3, \ldots, (k + 1)! + (k + 1)$ can be extended to include the composites $(k + 1)! + (k + 2), (k + 1)! + (k + 3), \ldots, (k + 1)! + (k + 2m)$, $(k + 1)! + (k + 2m + 1)$. $(k + 1)! + (k + 2m + 2)$ may, or may not, be a prime, depending on the value of k.*

Proof. Each element $k + 2j$, $1 \leqslant j \leqslant m$ of S has a prime factor q_j, where $q_j \leqslant \sqrt{k+2j}$. Since $\sqrt{k+2j} < k + 1$ for $k^2 + k > 2j - 1$, $q_j|(k + 1)!$, implying that $(k + 1)! + (k + 2j)$ is a composite.

When k is odd $(k \neq 1)$ such that $k + 2$ *is* a prime, we have, by Wilson's theorem, that $(k + 2 - 1)! \equiv -1 (\text{mod } k + 2)$. That is, $(k + 1)! + 1$ is divisible by $k + 2$ and thus is not a prime. Therefore, the consecutive composites $(k + 1)! + 2$, $(k + 1)! + 3, \ldots, (k + 1)! + (k + 1)$ can be extended (to the left) to include the composites $(k + 1)!$ and $(k + 1)! + 1$. $(k + 1)! - 1$ may, or may not, be a prime, depending on the value of k.

With regard to question 4 we note that for $k = 1$ we are asking whether the number of prime deserts of length 1 is finite or infinite. This is equivalent to asking if the number of twin primes is finite or infinite, a long standing unsolved problem in number theory.

Since the main purpose of this article is to consider consecutive prime deserts, in particular consecutive prime deserts each having the same length, we shall not pursue the above questions directly, but, rather, indirectly. (By n consecutive prime deserts we shall mean n deserts determined by $n + 1$ consecutive primes. Note that if the primes are also in arithmetic progression, the consecutive prime deserts all have the same length, and conversely. For a brief discussion of arithmetic progressions of primes see [**1**, pp. 10–15].) We begin with the following definition.

DEFINITION. Two consecutive prime deserts, each having the same length k, are called prime desert twins of length k. Three consecutive prime deserts, each having the same length k, are called prime desert triplets of length k. And so on. We may refer to n consecutive prime deserts, each having the same length k, as prime desert n-tuplets of length k.

Now suppose p is an odd prime, $p \neq 3$. To have desert twins of length k (where the deserts begin at $p + 1$) it must be the case that p, $p + (k + 1) = (p + 1) + k$, and $p + 2(k + 1) = (p + 2) + 2k$ are all prime integers. Since $p \neq 3$, one of $p + 1$, $p + 2$ must be divisible by 3. Thus $k \not\equiv 0 \pmod 3$ and we have $k \equiv 1 \pmod 3$ or $k \equiv 2 \pmod 3$. Suppose $k = 1 + 3m$. Then $p + 1 + k = (p + 2) + 3m$, implying $3 \nmid p + 2$, and $p + 2 + 2k = (p + 1) + 3(1 + 2m)$, implying $3 \nmid p + 1$. Therefore, it must be the case that $k \equiv 2 \pmod 3$. This, together with the fact that $k \equiv 1 \pmod 2$, tells us that $k \equiv 5 \pmod 6$. We have proved the following:

THEOREM 2. *Suppose p is an odd prime* $\neq 3$ *and k is odd. The existence of prime desert twins of length k implies that* $k \equiv 5 \pmod 6$.

So it is *necessary* to have $k \equiv 5 \pmod 6$ in order for prime desert twins of length k to exist. But is it the case that given such a k, there *are* prime desert twins of length k? Let us try to construct a desert twin of length $k = 5$. We will need to find integers $p, p + 1, p + 2, \ldots, p + 11, p + 12$ where p, $p + 6$, and $p + 12$ are primes and the remaining numbers are composites. Since p is a prime, each of $p + 1, p + 3, \ldots, p + 11$ is divisible by two. As noted above, assuming $p \neq 3$, either $3 | p + 1$ or $3 | p + 2$. Suppose the former. Then also $3 | p + 4$, $3 | p + 7$, and $3 | p + 10$. What about $p + 2$ and $p + 8$? They must also be composites or we will not have prime deserts of length 5. If we also assume $p \neq 5$, then one of the numbers $p + 1, p + 2, p + 3, p + 4$ must be divisible by 5. But since $p + 6$ and $p + 12$ must be primes, 5 cannot divide $p + 1$ or $p + 2$. Suppose $5 | p + 3$. Then also $5 | p + 8$. Finally, supposing $7 | p + 2$ gives us the situation described below:

n	0	1	2	3	4	5	6	7	8	9	10	11	12
divisors of $p + n$	–	2,3	7	2,5	3	2	–	2,3	5	2,7	3	2	–

Question: Is there a prime p such that $p + 6$ and $p + 12$ are also prime and such that the above is satisfied for the numbers $p + 1, \ldots, p + 5$ and $p + 7, \ldots, p + 11$? Answer: Yes. $p + 1$ divisible by 3, $p + 3$ divisible by 5, and $p + 2$ divisible by 7 suggests we need to find a prime p that satisfies the system of congruences

$$p \equiv 2 \pmod 3 \qquad p \equiv 2 \pmod 5 \qquad p \equiv 5 \pmod 7$$

where it is *also* true that $p + 6$ and $p + 12$ are primes. The Chinese remainder theorem guarantees an infinite number of solutions to such a system of linear congruences (all congruent to one another modulo $3 \cdot 5 \cdot 7$) but, of course, not all

solutions are primes. And when a particular solution p *is* a prime, $p + 6$ and/or $p + 12$ may not be prime.

It is routine to solve the above system and arrive at $p \equiv 47 \pmod{105}$. All solutions are of the form $p = 47 + 105m$ and we seek m such that p, $p + 6$, and $p + 12$ are primes. $m = 0$ leads to $p = 47$, $p + 6 = 53$, and $p + 12 = 59$, all primes. Thus 48, 49, 50, 51, 52 and 54, 55, 56, 57, 58 are prime desert twins of length $k = 5$. It is interesting to note that this desert twin is the *first* desert twin where $k = 5$.

Dirichlet's theorem tells us that every sequence of positive integers $\{a + bm\}$, where a and b are relatively prime, contains an infinite number of primes. Thus there is an infinite number of values of m for which $47 + 105m$ is a prime. (Note that m must be even.) Dirichlet's theorem also tells us that each of the sequences $\{53 + 105m\}$ and $\{59 + 105m\}$ contains an infinite number of primes. It would seem reasonable, then, to expect other values of m to produce three primes p, $p + 6$, $p + 12$, where $p \in \{105m + 47\}$, $p + 6 \in \{105m + 53\}$, and $p + 12 \in \{105m + 59\}$. The reader can check that this is the case for $m = 2$, 10, 50, and 74. But is the number of such m finite or infinite? Is the number of prime desert twins of length 5 finite or infinite?

One can certainly form a construction of the type used above for $k = 5$ to try and find a prime desert twin for any $k \equiv 5 \pmod 6$. Such a construction will lead to a system of congruences that will have a solution $p \equiv a \pmod t$ with $(a, t) = 1$ by the Chinese remainder theorem. This solution will lead to the three sequences $\{tm + a\}$, $\{tm + a + (k + 1)\}$, $\{tm + a + 2(k + 1)\}$ and each of these contains an infinite number of primes by Dirichlet's theorem. Each value of m that yields a prime in all three sequences gives a prime desert twin. *But*, will there *always* be such an m, and if so, is the number of such m finite or infinite?

Now let us turn our attention to prime desert triplets. Suppose p is a prime $\neq 3$. To have prime desert triplets beginning at $p + 1$ we must first have prime desert twins; thus $k \equiv 5 \pmod 6$ with $p + (k + 1)$ and $p + 2(k + 1)$ primes. For a desert triplet it also will have to be the case that $p + 3(k + 1)$ is a prime. Since $p \neq 3$ and $p + (k + 1)$ is a prime, $p + 3(k + 1)$ is not divisible by 3 or p. Suppose q is a prime > 3, $q \neq p$. If $q | (k + 1)$, then $q \nmid p + i(k + 1) \forall i$. If $q \nmid (k + 1)$, then for $i = 1, 2, 3$, we have $i(k + 1) \equiv a_i \pmod q$ where $1 \leqslant a_i \leqslant q - 1$ and $i \neq j \Rightarrow a_i \neq a_j$. Thus we may write

$$p + (k + 1) = (p + a_1) + qb_1$$
$$p + 2(k + 1) = (p + a_2) + qb_2$$
$$p + 3(k + 1) = (p + a_3) + qb_3.$$

Since p, $p + a_1$, $p + a_2$, and $p + a_3$ do not constitute the q consecutive integers from p to $p + q - 1$, $p + 3(k + 1)$ need not be divisible by q. Thus there are no further restrictions on k due to the prime q and so it is necessary only for k to be congruent to 5 modulo 6 to have prime desert triplets of length k.

Do prime desert triplets exist for every k such that $k \equiv 5 \pmod 6$? Can prime desert triplets of length k be constructed? If yes, then for $k = 5$ we would need to find primes p, $p + 6$, $p + 12$, and $p + 18$ such that the integers $p + 1, p + 2, \ldots, p + 5, p + 7, p + 8, \ldots, p + 11, p + 13, p + 14, \ldots, p + 17$ are all composites. Consider the following:

n	0	1	2	3	4	5	6	7	8	9	10	11	12
divisors of $p + n$	–	2, 3	7	2	3, 5	2	–	2, 3	11	2, 5, 7	3	2	–

n	13	14	15	16	17	18
divisors of $p + n$	2, 3	5	2	3, 7	2	–

This leads to the system of congruences

$$p \equiv 2 \pmod 3 \qquad p \equiv 1 \pmod 5 \qquad p \equiv 5 \pmod 7 \qquad p \equiv 3 \pmod{11}$$

which in turn leads to $p \equiv 971 \pmod{1155}$ or $p = 971 + 1155m$. Now we see that we need a value for m such that p, $p + 6$, $p + 12$, and $p + 18$ are all primes. The sequences $\{1155m + 971\}$, $\{1155m + 977\}$, $\{1155m + 983\}$, and $\{1155m + 989\}$ each contain an infinite number of primes, but is there a common value of m that yields a prime in each sequence? Yes; the first such m is $m = 210$. Thus we have constructed the prime desert triplets of length 5 determined by the four primes

243,521 243,527 243,533 243,539

Are there other values of m? How many?

The following construction leads to the first prime desert triplets of length $k = 5$. They are determined by the primes 251, 257, 263, and 269.

n	0	1	2	3	4	5	6	7	8	9	10	11	12
divisors of $p + n$	–	2, 3, 7	11	2	3, 5	2	–	2,3	7	2, 5	3	2	–

n	13	14	15	16	17	18
divisors of $p + n$	2, 3, 11	5	2, 7	3	2	–

So, at least in the case $k = 5$, it is actually possible to construct prime desert triplets.

One should soon wonder if it is possible to "extend" prime desert triplets to prime desert quadruplets. Well, if $k = 5$, we would need p, $p + 6$, $p + 12$, $p + 18$, *and* $p + 24$ all to be primes. If we assume $p \neq 5$, then either $5 | p + 1$, $5 | p + 2$, $5 | p + 3$, or $5 | p + 4$. Since $p + 6 = (p + 1) + 5$, $5 \nmid p + 1$ because otherwise no prime desert of length 5 would exist from p to $p + 6$. Since $p + 12 = (p + 2) + 10$, $5 \nmid p + 2$ because otherwise no prime desert twins would exist from p to $p + 12$. And since $p + 18 = (p + 3) + 15$, $5 \nmid p + 3$ else we would not have prime desert

triplets from p to $p + 18$. Therefore, it must be the case that $5 | p + 4$, implying that prime desert quadruplets of length 5 are *not* possible. A similar analysis will show that prime desert quadruplets of length $k = 11$ are not possible.

Since $29 \equiv 5 \pmod 6$, it is possible to have prime desert triplets of length $k = 29$. Might such a triplet be extended to a quadruplet? Unlike the cases above for $k = 5$ and $k = 11$, this time the answer is yes. If we have a prime desert triplet of length 29 beginning at $p + 1$ for some prime p, then we know that p, $p + 30$, $p + 60$, and $p + 90$ are primes. If it is also the case that $p + 120$ is a prime and $p + 91, p + 92, \dots, p + 119$ are composites, then we have a prime desert quadruplet. $p \neq 5$ implies that 5 does not divide any of the integers $p + 30$, $p + 60$, $p + 90$, or $p + 120$. Thus the restriction due to division by 5 that prevents the existence of prime desert quadruplets of length $k = 5$ and $k = 11$ (and $k = 17$ and $k = 23$ as well) does not hold for $k = 29$. In fact, prime desert quadruplets of length 29 *do* exist; the first such one is determined by the five primes

$$9{,}843{,}019 \qquad 9{,}843{,}049 \qquad 9{,}843{,}079 \qquad 9{,}843{,}109 \qquad 9{,}843{,}139$$

The above discussion regarding prime desert quadruplets suggests the following theorem:

THEOREM 3. *Suppose p is a prime, $p > 5$, and k is odd. The existence of prime desert quadruplets of length k implies that $k \equiv 29 \pmod{30}$.*

Proof. To have prime desert quadruplets of length k beginning at $p + 1$, it must be the case that $p, (p + 1) + k$, $(p + 2) + 2k$, $(p + 3) + 3k$, and $(p + 4) + 4k$ are all prime integers. Since $p > 5$, one of $p + 1$, $p + 2$, $p + 3$, $p + 4$ must be divisible by 5. Thus $k \not\equiv 0 \pmod 5$ and we have either that $k \equiv 1 \pmod 5$, $k \equiv 2 \pmod 5$, $k \equiv 3 \pmod 5$, or $k \equiv 4 \pmod 5$. Suppose $k = 1 + 5m$. Then $p + 1 + k = (p + 2) + 5m$, implying $5 \nmid p + 2$; $p + 2 + 2k = (p + 4) + 10m$, implying $5 \nmid p + 4$; $p + 3 + 3k = (p + 1) + 5(1 + 3m)$, implying $5 \nmid p + 1$; and $p + 4 + 4k = (p + 3) + 5(1 + 4m)$, implying $5 \nmid p + 3$. Therefore, $k \not\equiv 1 \pmod 5$. Similarly, it follows that $k \not\equiv 2 \pmod 5$, and $k \not\equiv 3 \pmod 5$. Thus it must be the case that $k \equiv 4 \pmod 5$. This, together with the fact that $k \equiv 5 \pmod 6$, implies $k \equiv 29 \pmod{30}$.

For $k = 29$ we can also note the following: $p + 30 = (p + 2) + 28$, $p + 60 = (p + 4) + 56$, $p + 90 = (p + 6) + 84$, $p + 120 = (p + 1) + 119$, $p + 150 = (p + 3) + 147$, and $p + 180 = (p + 5) + 175$. If $p \neq 7$, one of $p + 1, p + 2, \dots, p + 6$ must be divisible by 7. Thus if we have a prime desert quadruplet beginning at $p + 1$, 7 cannot be a factor of $p + 2$, $p + 4$, $p + 6$, or $p + 1$. If it is also the case that $7 \nmid p + 3$, a prime desert quintuplet is possible. In this event 7 must divide $p + 5$ and thus *at most* prime desert quintuplets are possible.

A necessary condition on the length k of prime desert n-tuplets is given by the following theorem, the proof of which is readily established by induction on i:

MAIN THEOREM. *Let p_i denote the ith odd prime and suppose k is odd. If $p > p_i$, then a necessary condition for the existence of prime desert $(p_i - 1)$-tuplets of length k*

beginning at $p + 1$ *is*

$$k \equiv (2 \cdot 3 \cdot 5 \cdots p_i - 1) \pmod{2 \cdot 3 \cdot 5 \cdots p_i}.$$

The same restriction on k *applies to prime desert* n*-tuplets for* n *such that* $p_i - 1 < n \leqslant p_{i+1} - 2$. *Thus if* $k \equiv (2 \cdot 3 \cdot 5 \cdots p_i - 1)(\bmod 2 \cdot 3 \cdot 5 \cdots p_i)$ *but* $k \not\equiv (2 \cdot 3 \cdot 5 \cdots p_{i+1} - 1)(\bmod 2 \cdot 3 \cdot 5 \cdots p_{i+1})$, *then a prime desert* $(p_i - 1)$*-tuplet may be extended to at most a prime desert* $(p_{i+1} - 2)$*-tuplet.*

REFERENCE

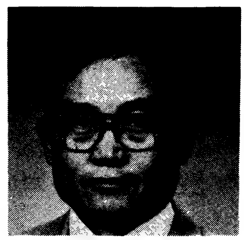
1. Richard K. Guy, Unsolved Problems in Number Theory, Springer-Verlag, New York, 1981.

The Determination of Kaprekar Convergence and Loop Convergence of All Three-Digit Numbers

KLAUS E. ELDRIDGE AND SEOK SAGONG

KLAUS E. ELDRIDGE did his undergraduate and graduate work in mathematics receiving a B.S. degree from Hardin-Simmons University, a M.S. degree from Oklahoma State University, and a Ph.D. degree from the University of Colorado (Boulder). Currently he is Chairman of the Computer Science Department at Ohio University, Athens.

SEOK SAGONG, born in Korea, received a B.S. from the State University of New York at Stony Brook and an M.S. from Ohio University, both in mathematics. Since 1982 he has been an assistant professor of Mathematics at Hardin-Simmons University, Abilene, TX.



Introduction. Take a four digit number x. Form a new number A by rearranging the digits of x in descending order. Form a new number B by rearranging the digits of x in ascending order. Define $K(x) = A - B$. In [**4**], D. R. Kaprekar discovered that after at most 7 such operations every four-digit number in base 10 leads to the same result, 6174.

In the late 60's and early 70's, the problem of proving and possibly generalizing the above discovery made the rounds at Berkeley, MIT, etc. ([**1**], [**6**]). D. D. McCracken said that the problem was called "the 6174 problem", and he conjectured that every three-digit number in an even base leads to a similar phenomenon [**5**].

The authors were introduced to the problem as such.

In this article, we study such an operation, called a Kaprekar transformation (transformation for short), on three-digit numbers in arbitrary bases. In these situations it is possible, as we shall see, to give quite a complete description of what happens:

(a) *For a three-digit number x in base r, the result of a single Kaprekar transformation is determined by the difference between the largest and smallest digits, I, of x and the base r: the first digit of the result is $I - 1$, the second digit $r - 1$, and the third digit $r - I$.*

If 728 is in base 15, for instance, then I is 6 and r is 15; hence, $(6 - 1, 15 - 1, 15 - 6)$, or $5\underline{14}9$ is the result after one transformation. Since I of $5\underline{14}9$ is 9, $8\underline{14}6$ is the result after the second transformation on 728. (We denote digits larger than 9 by representing the digits in base ten and underscoring. An ordered triple and a three-digit number with its digits written in juxtaposition will be used interchangeably, sometimes one preferred to the other with respect to clarity.)

(b) *There is a unique number, namely,* $((r - 2)/2, r - 1, r/2)$, *to which every three-digit number in base r converges after at most* $(r + 2)/2$ *transformations* (1 *transformation for* $r = 2$) *if and only if r is even.* (The unique number is called the Kaprekar constant.)

All three-digit numbers in base 1986, for instance, converge to the number $\underline{992}\ \underline{1985}\ \underline{993}$ after no more than 994 transformations.

(c) *There is a unique regenerative loop consisting of two numbers,* $((r - 3)/2, r - 1, (r + 1)/2)$ *and* $((r - 1)/2, r - 1, (r - 1)/2)$, *to which every three-digit number in base r leads after at most* $(r + 1)/2$ *transformations* (1 *for* $r = 3$) *if and only if r is odd.* (The two numbers, each the image of the other under the transformation, are called loop elements.)

In base 1987, for instance, every three-digit number becomes either $\underline{992}\ \underline{1986}\ \underline{994}$ or $\underline{993}\ \underline{1986}\ \underline{993}$ after at most 994 transformations.

(d) *The smallest number of transformations necessary for a three-digit number in an even base r to become the Kaprekar constant is* $2 + r/2 - I$ *if* $I < r/2$, *or* $I + 1 - r/2$ *if* $I \geq r/2$, *where I is as described in* (*a*).

Consider 867 in base 10, for instance. The difference I is 2 and r is 10. Hence, after exactly $2 + 10/2 - 2$, or 5 transformations it becomes the Kaprekar constant 495.

(e) *The smallest number of transformations necessary for a three-digit number in an odd base r to become a loop element is*: $(r + 3)/2 - I$ *if* $I < (r - 1)/2$; 1 *if* $I = (r - 1)/2$; $I - (r - 1)/2$ *if* $I > (r - 1)/2$.

If 867 is in base 15, for example, then I is 2 and r is 15. Hence, it requires at least $(15 + 3)/2 - 2$, or 7 transformations for 867 to become $7\underline{14}7$.

J. H. Jordan had proved a stronger version (even and only even bases) of McCracken's conjecture in his 1964 paper and identified Kaprekar constants in his proof [**2**, Theorem 3].

C. W. Trigg, 1974 paper, stated without proof the results mentioned in (b)–(e) above [**7**].

1. Preliminaries. We denote by *S(n, r)* the set of all positive integers expressible in n-digit sequences in base r exclusive of those digit sequences with all digits equal. For example, $S(3, 10)$ is the set of all decimal integers between 001 and 998

excluding 111, 222,..., 888. A "three-digit number" means an element of $S(3, r)$ for an appropriate base r.

Let $\boldsymbol{K}$: $S(n, r) \to S(n, r)$ be that transformation which permutes the n digits of a number in $S(n, r)$ to obtain the largest and the smallest number expressible using the same n digits and then subtracts the smallest from the largest. K is called the **Kaprekar transformation.** Recursively, we define

$$K^l(x) = \begin{cases} x & \text{for } l = 0 \\ K\left(K^{l-1}(x)\right) & \text{for } l = 1, 2, 3, \dots . \end{cases}$$

Let $x \in S(n, r)$. If $K(x) = x$, then x is said to be a **fixed point of** $\boldsymbol{K}$. If $x \in S(n,r)$ is a fixed point, and $K^m[S(n, r)] = \{x\}$ for some positive integer m, then x, if it exists, is called the **Kaprekar constant** of $S(n, r)$ and the smallest such integer m is called the **degree of Kaprekar convergence of** $\boldsymbol{S(n,r)}$. If $y \in S(n, r)$ and $K^l(y) = x$ for some positive integer l, where x is a fixed point of K in $S(n, r)$, then y is said to **converge to** $\boldsymbol{x}$ **under** $\boldsymbol{K}$ and the smallest such integer l is called the **degree of convergence of** $\boldsymbol{y}$ **to** $\boldsymbol{x}$. (Note that fixed points may not be unique; for instance, both 1001 and 0111 are fixed points in $S(4, 2)$.) If, in particular, x is the Kaprekar constant, then y is said to be **Kaprekar convergent** and l—the degree of convergence of y to x—is called the **degree of Kaprekar convergence of** $\boldsymbol{y}$.

If, for some integer m, $K^m(x) = K^{m+i}(x)$ and $K^m(x) \neq K^{m+j}(x)$ for all j such that $1 \leq j < i$, then $K^m(x)$ is said to **generate a loop of length** $\boldsymbol{i}$ and each of $K^{m+j}(x)$, $j = 0, 1, \dots, i - 1$, is called a **loop element.** (Note that being a unique fixed point does not necessarily mean the fixed point is the Kaprekar constant, and that the existence of a loop does not preclude the existence of a unique fixed point; in $S(4, 20)$, for example, $\underline{12}\ 3\ \underline{15}\ 8$ is a unique fixed point while $7\ \underline{19}\ \underline{19}\ \underline{12} \to \underline{12}\ 6\ \underline{12}\ 8 \to 6\ 3\ \underline{15}\ \underline{14} \to \underline{12}\ 7\ \underline{11}\ 8 \to 5\ 2\ \underline{16}\ \underline{15} \to \underline{14}\ 9\ 9\ 6 \to 7\ \underline{19}\ \underline{19}\ \underline{12}$ form a loop of length 6 with each element being the image of the preceding one under K.)

If $K^m(x)$ is a loop element for some integer m, then x is said to be **loop convergent** and the smallest such integer m is called the **degree of loop convergence of** $\boldsymbol{x}$. If every member of $S(n, r)$ is loop convergent, then $S(n, r)$ is said to be **loop convergent** and the smallest number of applications of K necessary for all members of $S(n, r)$ to be loop elements is called the **degree of loop convergence of** $\boldsymbol{S(n,r)}$.

We study the Kaprekar transformation on three-digit numbers only in this paper. Throughout the paper all numbers are nonnegative integers and K stands for the Kaprekar transformation; r for the base (radix) of a number system; I for the (positive) difference between the largest and smallest digits of a three-digit number.

For $x \in S(3\text{p}, r)$, we denote by ***d(n, i; x)*** the ith digit of $K^n(x)$ (counting from left to right) and by ***m(n, i; x)*** the ith digit in the ascending order of the magnitude of digits of $K^n(x)$ (that is, $m(n, 1; x) \leq m(n, 2; x) \leq m(n, 3; x)$ with at least one strict inequality). If there is no ambiguity, we simply write $d(n, i)$ and $m(n, i)$, respectively, for $d(n, i; x)$ and $m(n, i; x)$. For $x = 792$ in $S(3, 10)$, for example, $d(0, 1)$ is 7 while $m(0, 1)$ is 2, the smallest digit; $d(1, 1)$ is 6 while $m(1, 1)$ is 3 since $K(x) = 693$; $d(2, 1) = 5 = m(2, 2)$ and $d(2, 3) = 4 = m(2, 1)$ since $K^2(x) = 594$, and so on.

With these notations we can write the image of x in $S(3, r)$ after n applications of K as follows:

$$K^n(x) = (d(n,1;x),\, d(n,2;x),\, d(n,3;x))$$

or equivalently as $d(n,1;x)r^2 + d(n,2;x)r + d(n,3;x)$.

2. The mechanics of Kaprekar transformation. We begin with a fundamental lemma from which all our results will be derived.

LEMMA 1. *Let x be in $S(3, r)$. Then, for every $n \geq 1$, we have*

$$d(n,1;x) = m(n-1,3;x) - m(n-1,1;x) - 1 \tag{1}$$

$$d(n,2;x) = r - 1 \tag{2}$$

$$d(n,3;x) = r + m(n-1,1;x) - m(n-1,3;x). \tag{3}$$

Proof. Let $x \in S(3, r)$ and $n \geq 1$. By definition of the transformation K and due to the fact that the largest digit is greater than the smallest (hence, borrowing is necessary in the subtraction), we obtain

$$\begin{aligned} K^n(x) &= [m(n-1,3;x)r^2 + m(n-1,2;x)r + m(n-1,1;x)] \\ &\quad -[m(n-1,1;x)r^2 + m(n-1,2;x)r + m(n-1,3;x)] \\ &= [m(n-1,3;x) - m(n-1,1;x) - 1]r^2 \\ &\quad +[r + m(n-1,2;x) - m(n-1,2;x) - 1]r \\ &\quad +[r + m(n-1,1;x) - m(n-1,3;x)]. \end{aligned}$$

Hence, the first digit of $K^n(x)$ is $m(n-1,3) - m(n-1,1) - 1$, the second is $r - 1$, and the third is $r + m(n-1,1) - m(n-1,3)$. ∎

From Lemma 1, one can readily see that, for every $n \geq 1$,

$$\begin{aligned} d(n,1;x) + d(n,3;x) = d(n,2;x) &= r - 1 \\ &= m(n,3;x) = m(n,1;x) + m(n,2;x). \end{aligned} \tag{4}$$

From (1) and (4), it follows that, for all $n \geq 1$,

$$d(n+1,1;x) = m(n,2;x) - 1; \tag{5}$$

and from (3) and (4)

$$d(n+1,3;x) = m(n,1;x) + 1. \tag{6}$$

The lemma also shows that the image $K(x)$ of a three-digit number x in a particular base r is determined by the (positive) difference between the largest and smallest digits of x; hence, one needs to know only this difference to evaluate $K(x)$. We denote this difference by $I(n;x)$: for $x \in S(3, r)$,

$$\boldsymbol{I(n;x) = m(n,3;x) - m(n,1;x)} \quad \text{for every } n \geq 0.$$

Note we use I for $I(0, x)$ and that: if x and y are in $S(3, r)$, then for each $n \geq 1$,

$$K^n(x) = K^n(y) \quad \text{if and only if} \quad I(n-1;x) = I(n-1;y).$$

Though K may be thought of as a function of I, after the initial application of K the middle digit, being $r-1$, remains always the largest. So we investigate the behavior of the first and third digits under K, which are, in some order, the smallest and middle-sized digits. We denote by $J(n;x)$ the (nonnegative) difference between the smallest and middle-sized digits; that is, for $x \in S(3,r)$,

$$\boldsymbol{J(n;x) = m(n,2;x) - m(n,1;x)} \quad \text{for } n \geq 0.$$

Then, for $n \geq 1$,

$$J(n;x) = m(n,2;x) - m(n,1;x) = |d(n,1;x) - d(n,3;x)|.$$

If we write (1) and (3) using $I(n-1;x)$ for $m(n-1,3;x) - m(n-1,1;x)$, then $J(n;x) = |r - 2I(n-1;x) + 1|$. Thus, for $n \geq 1$,

$$J(n;x) = \begin{cases} r - 2I(n-1;x) + 1 & \text{if } I(n-1;x) \leq (r+1)/2 \\ 2I(n-1;x) - r - 1 & \text{if } I(n-1;x) \geq (r+1)/2 \end{cases}. \tag{7}$$

Lemma 2. *Let x be in $S(3,r)$. Then, for all $n \geq 1$, $J(n,x)$—the (nonnegative) difference between the first and third digits of $K^n(x)$—is odd if and only if the base r is even.*

Proof. By (4), the base r and $m(n,1) + m(n,2)$ have opposite parity. But, $m(n,2) - m(n,1)$ *and* $m(n,1) + m(n,2)$ have the same parity. ■

From (5) and (6), it follows that, for $n \geq 1$,

$$d(n+1,1;x) - d(n+1,3;x) = J(n;x) - 2.$$

Hence, for every $n \geq 1$ and every x in $S(3,r)$,

$$d(n+1,3;x) \leq d(n+1,1;x) \quad \text{if and only if} \quad J(n;x) \geq 2. \tag{8}$$

Consequently, if $J(n;x) \geq 2$ for any $n \geq 1$, then by the definition of J and equations (5) and (6),

$$J(n+1;x) = J(n;x) - 2.$$

If $J(1;x) \geq 2$, therefore, then each application of K decreases $J(1;x)$ by 2 as long as it is not already less than 2. Hence, $J(1+i;x) = J(1;x) - 2i$ for all $i \leq J(1;x)/2$. Thus we have proved

Lemma 3. *Let x be in $S(3,r)$. If $J(1;x) \geq 2$, then*

$$J(1+i;x) = J(1;x) - 2i \text{ for all } i \leq \lfloor J(1;x)/2 \rfloor.$$

3. Determination of Kaprekar convergence and loop convergence. According to Lemma 1, for every x in $S(3,r)$,

$$K(x) = (I-1, r-1, r-I). \tag{9}$$

Hence, $K[S(3,r)]$ has exactly $r-1$ elements, namely, $(I-1, r-1, r-I)$, $I = 1,2,\ldots,r-1$; moreover, for every $n \geq 1$, $K^n(x)$ must be one of these $r-1$ numbers. For some $n \leq r$, therefore, $K^n(x)$ must be either a fixed point or a loop

element that starts a regenerative loop. It turns out that whether the base is even or odd determines which of these alternatives actually happens.

Lemma 3 implies that $J(1; x)$ decreases by 2 each time K is applied, eventually reaching 1 or 0 according as the base r is even or odd (due to Lemma 2), yet says nothing about the values of $J(\ ; x)$ thereafter. By (5) and (6), however,

$$J(n+1; x) = |J(n;x) - 2| \quad \text{for every } n \geq 1.$$

Therefore, $J(n; x) = 1$ for all $n \geq l$ if $J(l; x) = 1$ for some $l \geq 1$ whereas $J(n; x)$ is alternately 0 and 2 for all $n \geq l$ if $J(l; x) = 0$ for some $l \geq 1$. So the difference between the first and third digits of a fixed point, if it exists, is 1 whereas that of a loop element, if it exists, is either 0 or 2.

Now, let $x \in S(3, r)$ and write $k = J(1; x)$.

Suppose k is odd. (Note that this happens if and only if r is even, by Lemma 2.) If $k > 1$, then, by Lemma 3, $J(l; x) = 1$ for some $l \geq 2$ and the smallest such integer l is $(k+1)/2$; hence, by (8), $d((k+1)/2, 3; x) + 1 = d((k+1)/2, 1; x)$. Applying (4) to this and using (5) and (6), we see that $K^{(k+3)/2}(x)$ is $((r-2)/2, r-1, r/2)$ and is a fixed point of K. Since $k > 1$ if and only if I is neither $r/2$ nor $(r+2)/2$, $(k+3)/2$ is

$$r/2 + 2 - I \ (\text{if } I < r/2) \quad \text{or} \quad I - r/2 + 1 \ (\text{if } I > (r+2)/2)$$

by (7).

If $k = 1$, there are two cases to consider: $I = r/2$ and $I = (r+2)/2$. If $I = r/2$, then, by Lemma 1, $K(x)$ is equal to $((r-2)/2, r-1, r/2)$; if $I = (r+2)/2$, then $K^2(x)$ is equal to $((r-2)/2, r-1, r/2)$.

Thus, $((r-2)/2, r-1, r/2)$ is the fixed point to which every member of $S(3, r)$ converges and hence, it is the Kaprekar constant if and only if r is even. Moreover, the smallest number of transformations K necessary for a three-digit number x in $S(3, r)$ to converge to the Kaprekar constant is

$$\begin{aligned} &2 - I + r/2 \quad \text{if } I < r/2, \\ &1 + I - r/2 \quad \text{if } I \geq r/2. \end{aligned}$$

Next, suppose that k is even (this happens if and only if r is odd). If $k \geq 2$, then, by Lemma 3, $J(l; x) = 0$ for some $l \geq 2$ and the smallest such l is $(k+2)/2$. Applying (4), we see that $K^{(k+2)/2}(x)$ is $((r-1)/2, r-1, (r-1)/2)$. And applying K twice more, we can show that $((r-1)/2, r-1, (r-1)/2)$ and $((r-3)/2, r-1, (r+1)/2)$ form a regenerative loop of length 2 as each is the image of the other under K. (Note the values of $J(\ ; \ )$ of these numbers are 0 and 2, respectively; and $J(k/2; x)$ is also 2, but $K^{k/2}(x)$ is not a loop element.)

If $k = 0$, then by (4), $K(x) = ((r-1)/2, r-1, (r-1)/2)$, a loop element.

Thus, $\cdots \to ((r-1)/2, r-1, (r-1)/2) \to ((r-3)/2, r-1, (r+1)/2) \to ((r-1)/2, r-1, (r-1)/2) \to \cdots$ form a regenerative loop of length 2 to which every member of $S(3, r)$ converges if and only if r is odd.

From Lemma 1, we see that

$$K(x) = ((r-1)/2, r-1, (r-1)/2) \quad \text{if and only if} \quad I = (r+1)/2;$$

whereas

$$K(x) = ((r-3)/2, r-1, (r+1)/2) \quad \text{if and only if} \quad I = (r-1)/2.$$

If I is neither $(r+1)/2$ nor $(r-1)/2$, then, by (7), $k \geq 2$ with the equality holding only if $I = (r+3)/2$, and $(k+2)/2$—the smallest number of transformations K necessary for x to become a loop element in this case—is, by (7),

$$(r+3)/2 - I \ (\text{if } I < (r-1)/2) \quad \text{or} \quad I - (r-1)/2 \ (\text{if } I > (r+1)/2).$$

Accordingly, the smallest number of transformations K necessary for a three-digit number x in $S(3, r)$ to become a loop element is

$$\begin{array}{cl} (r+3)/2 - I & \text{if } I < (r-1)/2, \\ 1 & \text{if } I = (r-1)/2, \\ I - (r-1)/2 & \text{if } I \geq (r+1)/2. \end{array}$$

We have thus proved the main result, namely,

THEOREM 1.

(A) *The set* $S(3, r)$ *has the Kaprekar constant* $((r-2)/2, r-1, r/2)$ *if and only if* r *is even.*

(B) *The set* $S(3, r)$ *is loop convergent with a unique regenerative loop of length* 2 *consisting of* $((r-3)/2, r-1, (r+1)/2)$ *and* $((r-1)/2, r-1, (r-1)/2)$ *if and only if* r *is odd.*

(C) *If* x *in* $S(3, r)$, *where* r *is even, is not the Kaprekar constant, then its degree of Kaprekar convergence is given by*:

(i) $2 - I + r/2$ if $I < r/2$

(ii) $1 + I - r/2$ if $I \geq r/2$.

(D) *If* x *in* $S(3, r)$, *where* r *is odd, is not a loop element, then its degree of loop convergence is given by*:

(i) $(r+3)/2 - I$ *if* $I < (r-1)/2$

(ii) 1 *if* $I = (r-1)/2$

(iii) $I - (r-1)/2$ *if* $I > (r-1)/2$.

COROLLARY 1.

(A) *The degree of Kaprekar convergence of* $S(3, r)$, *where* r *is even, is* $1 + r/2$ *for* $r \geq 4$ *and* 1 *if* $r = 2$.

(B) *The degree of loop convergence of* $S(3, r)$, *where* r *is odd, is* $(r+1)/2$ *for* $r \geq 5$ *and* 1 *if* $r = 3$.

Proof. The degree of convergence of the set $S(3, r)$ is equal to the maximum degree of convergence over all members of $S(3, r)$. Hence, from (C) and (D) of Theorem 1 follows the corollary. ■

4. Additional examples. We combine (5) and (6) into

$$K^n(x) = (m(n-1, 2; x) - 1, r-1, m(n-1, 1; x) + 1) \text{ for } n \geq 2. \quad (10)$$

Then, we can use (9) and (10) as an algorithm to evaluate $K^n(x)$ for all $n \geq 1$.

Consider 787, for example. If it is in base 10, then using (9) and (10) we easily obtain the successive images: $787 \to 099 \to 891 \to 792 \to 693 \to 594 \to 495 \to 495 \to \cdots$; if it is in base 15, its successive images are: $787 \to 0\ \underline{14}\ \underline{14} \to \underline{13}\ \underline{14}\ 1 \to \underline{12}\ \underline{14}\ 2 \to \underline{11}\ \underline{14}\ 3 \to \underline{10}\ \underline{14}\ 4 \to 9\ \underline{14}\ 5 \to 8\ \underline{14}\ 6 \to 7\ \underline{14}\ 7 \to 6\ \underline{14}\ 8 \to 7\ \underline{14}\ 7 \to 6\ \underline{14}\ 8 \to \cdots$.

Examples like these, which can be readily obtained (by a computer or mentally), can be used to substantiate the formulas in Theorem 1 and its corollary as well as some properties of K described elsewhere in the paper. Notice that we chose the number 787 so as to exhibit an element of $S(3, r)$ whose degree of convergence coincides with that of the set $S(3, r)$. ($K^n[S(3,10)] = \{495\}$ if and only if $n \geq 6$, and $Kn[S(3,15)] = \{7\ \underline{14}\ 7,\ 6\ \underline{14}\ 8\}$ if and only if $n \geq 8$.)

The authors would like to mention, in passing, that the material presented in this paper provides answers to simple questions—perhaps suitable for exercise problems for elementary courses—such as:

1. How many three-digit numbers in base 10 will become 892 under K? 297 under K?
2. Given a base, say 1986, determine the number of all three-digit numbers in $S(3,1986)$ whose degree of Kaprekar convergence is exactly that of $S(3,1986)$.
3. If a three-digit number is randomly chosen in base 10, what is the probability that the number will become 495 after 2 Kaprekar transformations? After 5 transformations?

REFERENCES

1. Elementary Problems (E2222), this MONTHLY, 77(1970) 307.
2. J. H. Jordan, Self Producing Sequences of Digits, this MONTHLY, 71(1964) 61–64.
3. D. R. Kaprekar, Another Solitaire Game, Scripta Math. 15(1949) 244–245.
4. ______, An Interesting Property of the Number 6174, Scripta Math., 21(1955) 304.
5. D. D. McCracken, A Guide to Fortran IV Programming, 2nd ed., John Wiley & Sons, Inc., New York, 1972, 252.
6. Solutions of Elementary Problems (Kaprekar's constant), this MONTHLY, 78(1971) 197–198.
7. C. W. Trigg, All Three-Digit Integers Lead to..., The Mathematics Teacher, 67(1974) 41–45.

The Editor's Corner: The Quest for Normality

HERBERT S. WILF

The integral. Until now these columns have discussed research developments. This month the offering is more in the spirit of the 'Teaching of Mathematics' department: the discussion will concern a subject that has been well known to specialists for years, but deserves wider dissemination. We are going to do two things. First we will evaluate a certain n-fold definite integral. Second, although the evaluation is pretty enough to need no further justification, it has one: from it we can (and will) derive the form of the normal distribution function in n variables.

The integral that you are about to see is available in various references, the application of it is also to be found in standard places (e.g. [**1**]), but hopefully some readers will find a new and interesting addition to a linear algebra course somewhere in the paragraphs below. It is a relative of the well-known evaluation

$$\int_{-\infty}^{\infty} e^{-t^2}\,dt = \sqrt{\pi}\,. \tag{1}$$

All of us who have taught calculus know the devices that combine to prove (1). We write down the integral twice, using different dummy variables, push together the two integrals, switch to polar coordinates, etc., etc. My students always think that they're learning another general technique for integration. I try to tell them that there are virtually no other applications of this whole chain of events, but I don't think they believe me. The related integral is given by the following theorem.

THEOREM 1. *Let A be an $n \times n$ real, symmetric, positive-definite matrix. Then*

$$\underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{n \text{ times}} e^{-(\mathbf{x},A\mathbf{x})}\,dx_1\,dx_2\,\cdots\,dx_n = \frac{\pi^{n/2}}{\sqrt{\det(A)}}\,. \tag{2}$$

In the above, $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, and $(\mathbf{a},\mathbf{b})$ is the usual inner product of two vectors $\mathbf{a}$ and $\mathbf{b}$.

Before we get down to the proof, which is the really nice part, here is a small example. Suppose you wanted to calculate

$$I = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-5x^2-6y^2+4xy}\,dx\,dy. \tag{3}$$

First notice that the argument of the exponential is as in (2), where the matrix A is

$$A = \begin{pmatrix} 5 & -2 \\ -2 & 6 \end{pmatrix}.$$

A quick check shows that A is positive definite, and from (2) we find painlessly that

$$I = \frac{\pi}{\sqrt{26}}\,.$$

Now let's prove the theorem. First, since A is real and symmetric, we can write $A = P\Lambda P^T$, where P is a real orthogonal matrix, Λ is $\operatorname{diag}(\lambda_1, \dots, \lambda_n)$, and the λ's are the eigenvalues of A. Make a change of variables in (2) by letting $y = P^T x$. Orthogonal matrices being what they are, this substitution is inverted by $x = Py$.

What happens to the argument of the exponential on the left side of (2) is that

$$\begin{aligned}(x, Ax) &= (Py, APy) \\ &= (Py, P\Lambda P^T P y) \\ &= (y, P^T P \Lambda P^T P y) \\ &= (y, \Lambda y) \\ &= \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2,\end{aligned} \tag{4}$$

which is to say that the quadratic form has (scarcely unexpectedly) been reduced to a sum of squares.

Further, since P is an orthogonal matrix, the volume element $dx_1\, dx_2 \cdots dx_n$ in the integral is simply replaced by $dy_1\, dy_2 \cdots dy_n$.

Where we are at the moment is that the integral on the left side of (2) has become, after this rotation of coordinates,

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-\lambda_1 y_1^2 - \cdots - \lambda_n y_n^2}\, dy_1\, dy_2 \cdots dy_n. \tag{5}$$

We pause briefly to express appreciation for the positivity of the λ's, and then we observe that (5) is just a product of n one-dimensional integrals, *viz.*

$$\left\{\int_{-\infty}^{\infty} e^{-\lambda_1 y_1^2}\, dy_1\right\} \cdots \left\{\int_{-\infty}^{\infty} e^{-\lambda_n y_n^2}\, dy_n\right\}.$$

Each of these is trivially reducible to (1) above, and the left side of (2) has now been found to be equal to

$$\sqrt{\frac{\pi}{\lambda_1}}\sqrt{\frac{\pi}{\lambda_2}} \cdots \sqrt{\frac{\pi}{\lambda_n}}.$$

The proof of the theorem is completed by noting that the product of the eigenvalues of a matrix is equal to its determinant. ■

If you liked that one, or maybe knew it already, then try this generalization.

THEOREM 2. *Let ω be a given vector, and let A be an $n \times n$ real, symmetric, positive definite matrix. Then*

$$\int e^{-i(\omega, x)} e^{-1/2(x, Ax)}\, dx = \frac{(2\pi)^{n/2}}{\sqrt{\det A}} e^{-1/2(\omega, A^{-1}\omega)}, \tag{6}$$

where '$\int$' means the integral over all of $\mathbf{E}^n$.

Equation (6) is, of course, the calculation of a Fourier transform.

The appearance of the inverse matrix, A^{-1}, on the right side of (6) is quite extraordinary, as is the symmetry between the function whose transform is being

taken and the shape of the transform itself. If you haven't ever had the experience of seeing this inverse appear before your very eyes then by all means do try the derivation. It begins in the same way as the proof of theorem 1, but after the rotation of the coordinate axes and the evaluation of the integral, the thing to do is not to stop there but to relate the mysterious looking sums that will appear to the inverse matrix of A. Be sure to remember that since $A = P\Lambda P^T$ it must be true that

$$\sum_k (P^T)_{k,r}(P^T)_{k,s}/\lambda_k = (A^{-1})_{r,s} \qquad (r, s = 1, n),$$

and that is really all you'll need to prove (6).

The 'bell-shaped curve' in *n*-space. My motives in discussing the integrals (2), (6) were not entirely pure (as opposed to 'applied'). I also wanted to share a fruitful application of those integrals with you, namely to the generalization of the 'bell-shaped curve' to higher dimensional space.

In one dimension we have the familiar 'normal curve,' given by

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(1/2)((x-\mu)/\sigma)^2}. \tag{7}$$

Furthermore, in (7), we know the 'meanings' of the parameters μ and σ: μ is the average value of x and σ^2 is the average value of $(x - \mu)^2$.

Today's question is: what shall we replace (7) with when we get to n variables $\mathbf{x} = (x_1, x_2, \ldots, x_n)$? In other words, what is the form of the n-variable normal distribution, and what are the meanings of the parameters that it contains?

To begin answering this question, let A be an $n \times n$ real, symmetric, positive definite matrix, and let

$$F(x) = \frac{\sqrt{\det A}}{(2\pi)^{n/2}} e^{-(1/2)(x, Ax)}. \tag{8}$$

According to (2) above, the integral of F over all space is 1. Since F is nonnegative-valued, it follows that F is a *probability density function.*

Next, in terms of this F, the Fourier transform relation (6) takes the form

$$\int e^{-i(\omega, x)} F(x)\, dx = e^{-(1/2)(\omega, A^{-1}\omega)}. \tag{9}$$

The thing to do now is to expand both sides of (9) in a power series in ω and compare the coefficients. If we do this we find that the left-hand side of (9) is

$$1 - i\sum_{r=1}^{n} \omega_r \int x_r F(x)\, dx - \frac{1}{2}\sum_{r,s} \omega_r\omega_s \int x_r x_s F(x)\, dx + \cdots, \tag{10}$$

whereas the right-hand side is

$$1 - \frac{1}{2}\sum_{r,s} (A^{-1})_{r,s}\omega_r\omega_s + \cdots. \tag{11}$$

If we equate the coefficients of ω_r on both sides we discover the unsurprising fact that

$$\int x_r F(x)\,dx = 0 \qquad (r = 1, 2, \ldots, n). \tag{12}$$

If we equate the coefficients of $\omega_r \omega_s$ on both sides we discover the possibly surprising fact that

$$\int x_r x_s F(x)\,dx = (A^{-1})_{r,s} \qquad (r, s = 1, 2, \ldots, n). \tag{13}$$

The conclusion? *With respect to the probability density function $F(x)$ of (8), the average value of each of the coordinates x_r is 0, and the average value of each of the products $x_r x_s$ is $(A^{-1})_{r,s}$.*

Probabilists refer to the average of $x_r x_s$ as the *covariance* of the two variables x_r and x_s. Hence we have shown that A^{-1} is the matrix of covariances. Since that is the case, why don't we change its name? Instead of A, let's call our matrix C^{-1}. Then C will be the covariance matrix. One final touch would be to allow for the possibility that the variables have mean values other than 0. If μ_i is the mean value of the ith variable and $\boldsymbol{\mu}$ is the vector of all of the μ_i, then we can summarize our findings in the form of

THEOREM 3. *Let $\boldsymbol{\mu}$ be a given n-vector of constants, and let C be a given positive definite $n \times n$ matrix. Then the function*

$$F(x) = \frac{1}{(2\pi)^{n/2}\sqrt{\det C}} \exp\left\{ -\frac{1}{2}\left((\mathbf{x} - \boldsymbol{\mu}), C^{-1}(\mathbf{x} - \boldsymbol{\mu})\right)\right\} \tag{14}$$

is a probability density function (the 'normal' distribution) in $\mathbf{E}^n$. With respect to this density, the average value of the rth coordinate x_r is μ_r $(r = 1, \ldots, n)$, and the average value of $(x_r - \mu_r)(x_s - \mu_s)$ is $C_{r,s}$ $(r, s = 1, \ldots, n)$.

Do take a minute to let $n = 1$ in (14) and compare with (7)!

REFERENCE

1. Harald Cramér, Mathematical Methods of Statistics, Princeton, 1974.

NOTES

EDITED BY DENNIS DETURCK, RICHARD LIBERA, AND ANITA E. SOLOW

What Pythagoras Could Have Done

YORAM SAGHER
Department of Mathematics, Statistics, and Computer Science, University of Illinois, Chicago

Pythagoras allegedly proved that the square root of 2 is irrational. Plato's dialogue *Theaitetos* tells of proofs by Theodorus of Kyrene of the irrationality of the square roots of all integers less than or equal to 17, with the obvious exceptions. Why Theodorus stopped at 17 is a subject of lively speculation, but the general tone of modern writers (see, e.g., [**1**]) is to sympathize with the giants of the classical age for not having had access to the properties of prime numbers. Had they but known that p divides a^2 implies p divides a, they would have seen through the problem at a glance.

In an issue of the MONTHLY [**2**], W. C. Waterhouse discusses a plausibility argument for the square root of k being either an irrational or an integer, based on the assumption that the properties of prime numbers are as mysterious to our students as they were to the Greeks. However, the plausibility argument is but a rewording of the usual idea.

In preparing a talk to high school students, the following proof occurred to me. The main point of the proof is that it does not depend on properties of prime numbers, and so was fully accessible to Pythagoras and to Theodorus. It is also accessible to high school students after one year of algebra. When shown to a number of bright ninth-graders, it caused some excitement.

Suppose $\sqrt{k} = m/n$, where m and n are integers with $n > 0$. If k is not a square, there exists an integer q so that $q < m/n < q + 1$. Now $m^2 = kn^2$ implies $m(m - qn) = n(kn - qm)$ and, hence, $m/n = (kn - qm)/(m - qn)$. From $q < m/n < q + 1$ we get $0 < m - qn < n$. Therefore we have:

$$\sqrt{k} = (kn - qm)/(m - qn),$$

where the denominator is positive and smaller than the one in the original fraction. Continuing, we get an infinite decreasing sequence of positive integers, an impossibility.

REFERENCES

1. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, Oxford University Press, Oxford, 1960.
2. W. C. Waterhouse, Why Square Roots are Irrational, *American Mathematical Monthly*, 93 (1986) pp. 213–214.

A Question of Protocol

I. BOUWER* and Z. STAR
Department of Mathematics and Statistics, University of New Brunswick, Fredericton, Canada E3B 5A3.

1. Introduction. Relative to any ordering of the n vertices of the path graph $\pi(n)$, $n \geqslant 2$, the greedy coloring algorithm produces a 2-coloring or a 3-coloring. We determine the number of vertex orderings that result in 2-colorings. We also solve the corresponding problem for the cycle graph $\gamma(n)$ on n vertices, $n \geqslant 3$.

Let G be any finite simple graph, and S any ordered set whose elements will be called *colors*. (For graph-theoretic terms not defined below, we refer the reader to **[1]**.) For any given ordering Γ of the vertices of G, the *greedy coloring of G relative to* Γ is the coloring of G found by considering the vertices of G in turn, in the order provided by Γ, and assigning to each vertex the lowest possible color, relative to the order on S, such that no two adjacent vertices are assigned the same color. Note that if the maximum degree of the vertices of G is equal to d, then any given ordering of the vertices of G results in a greedy coloring with at most $d+1$ colors. An *r-color vertex ordering* of G is an ordering of the vertices of G relative to which the greedy coloring of G uses exactly r colors. If G has n vertices, then a vertex ordering of G will be indicated by a one-to-one assignment of the numbers $1, 2, \ldots, n$ to the vertices of G.

For any natural number $n \geqslant 2$, let $\pi(n)$ denote the labeled *path graph* on n vertices (see Figure 1).

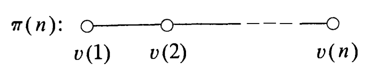


FIG. 1

The labeled *cycle graph* on n vertices, $\gamma(n)$, is the labeled graph consisting of the vertices and edges of an n-gon. Since the vertices of either graph have maximum degree equal to 2, it follows that any vertex ordering results in a greedy coloring with just two or three colors.

An interesting and unexpectedly difficult problem (raised in **[2]**) is to determine the number of 2-color vertex orderings $P(n)$ of $\pi(n)$, and $C(n)$ of $\gamma(n)$. The problem may be envisaged as follows: The chairs in a royal banquet hall are arranged along one side of a rectangular table (or around a circular table with an even number of chairs), and colored alternately red and blue. The king arrives first and takes any seat. The other members of the court arrive in descending order of rank and can sit anywhere, except that protocol requires a person to be seated in the

*Supported by NSERC grant A7332.

same colored seat as the king or else next to somebody already seated. If there are n persons altogether, and n seats, what is the number of possible seatings?

We shall let $O(n)$ (or $E(n)$) denote the number of 2-color vertex orderings of $\pi(n)$ in which the number 1 is assigned to a vertex in an odd (or even) position. Thus, $O(n) + E(n) = P(n)$. In Theorem 1, we derive simple recurrence relations for $O(n)$, and, in Section 3, explicit solutions for $O(n)$, and thence for $P(n)$ and $C(n)$ (see Proposition 1), generating functions yielding in terms of Bernoulli numbers.

2. Recurrence relations. Parts (i) and (ii) of the following Proposition appear in [**2**]; however, the proof given here is shorter.

PROPOSITION 1. *For* $n \geqslant 1$:
(i) $P(2n) = 2O(2n)$,
(ii) $P(2n+1) = 2n(2n+1)O(2n-1)$,
(iii) $C(2n+2) = (2n+2)P(2n+1)$, *and*
(iv) $C(2n+1) = 0$.

Proof. Note that $O(2n) = E(2n)$, since an odd position when counted from the left is an even position when counted from the right. This implies (i). (iv) is immediate. To prove (iii), we note that in a vertex ordering of $\gamma(2n+2)$, any one of the $2n+2$ vertices can carry the label $2n+2$, while the first $2n+1$ labels determine a 2-color vertex ordering of $\pi(2n+1)$. To prove (ii), we let two vertex orderings of $\gamma(2n+2)$ be equivalent if one can be derived from the other by a rotation. One then obtains a bijection between the set of equivalence classes of 2-color vertex orderings of $\gamma(2n+2)$ and the set of 2-color vertex orderings of $\pi(2n+1)$ by deleting the vertex in the cycle that is assigned the number $2n+2$, and taking the remaining vertices in clockwise order, say. But in any vertex ordering of the cycle, the two vertices on either side of the vertex to which the number 1 is assigned may be labeled in $2n(2n+1)$ ways, and the number of those orderings of the remaining $2n-1$ vertices that result in a greedy 2-coloring of $\gamma(2n+2)$ is equal to $O(2n-1)$. ■

THEOREM 1. (i) *For* $n \geqslant 2$,

$$O(2n+1) = 4nO(2n-1) + 4\sum_{k=1}^{n-1}\binom{2n}{2k}k(n-k)O(2k-1)O(2n-2k-1)$$

and (ii) *for* $n \geqslant 3$,

$$\begin{aligned}O(2n) &= (2n-1)O(2n-2) + (2n-1)(2n-2)O(2n-3)\\ &\quad + 2\sum_{k=1}^{n-2}\binom{2n-1}{2k}k(2n-2k-1)O(2k-1)O(2n-2k-2).\end{aligned}$$

Proof. We prove (i); (ii) follows similarly. The determination of $O(2n+1)$ reduces to two cases. In the first case, the number 1 is assigned to an end-vertex $v(1)$ or $v(2n+1)$. When 1 is assigned to $v(1)$, the vertex $v(2)$ can be assigned any one of the $2n$ remaining numbers, and, in each instance, the number of orderings of

$v(3), \ldots, v(2n+1)$ that result in 2-colorings is $O(2n-1)$. By the reflective symmetry of $\pi(n)$, we arrive at a similar conclusion when 1 is assigned to $v(2n+1)$. This case then accounts for the term $4nO(2n-1)$. In the second case, 1 is assigned to an interior vertex $v(2k+1)$,

$$1 \leqslant k \leqslant n-1.$$

The set of $2k$ numbers assigned to the first $2k$ vertices $v(1), \ldots, v(2k)$ may be selected in

$$\binom{2n}{2k}$$

ways. For each such selection, $v(2k)$ may be assigned any one of the $2k$ numbers selected, and the number of orderings of the first $2k-1$ vertices resulting in 2-colorings of $\pi(n)$ is then $O(2k-1)$. The vertex $v(2k+2)$ may be assigned any one of the remaining $2n-2k$ numbers, and then the number of orderings of $v(2k+3), \ldots, v(2n+1)$ resulting in 2-colorings of $\pi(n)$ is equal to $O(2n-2k-1)$. ■

Since, in a 2-color vertex ordering of $\pi(n)$ or $\gamma(n)$, any neighbor of the vertex assigned 1 may be assigned any one of $n-1$ labels, we remark that $n-1$ divides each of $E(n)$, $O(n)$, $P(n)$, $C(n)$.

The following table gives the values of $O(n)$, $P(n)$, and $C(n)$, for $n \leqslant 10$:

n	$O(n)$	$P(n)$	$C(n)$	$n!$
1	1	1	—	1
2	1	2	—	2
3	4	6	0	6
4	9	18	24	24
5	56	80	0	120
6	185	370	480	720
7	1632	2352	0	5040
8	7217	14434	18816	40320
9	81664	117504	0	362880
10	451089	902178	1175040	3628800

3. Explicit solution. It is easy to obtain explicit generating functions for these sequences. Let

$$G(t) = O(1)t/1! + O(3)t^3/3! + \cdots .$$

Multiply (i) of Theorem 1 by $t^{2n}/(2n)!$ and sum from 2 to ∞, to obtain

$$G'(t) = (tG(t)+1)^2. \tag{1}$$

The substitution

$$y = -1 + t + 1/G(t)$$

reduces (1) to the Bernoulli type differential equation

$$y' + 2y + y^2 = 0,$$

with solution:

$$y = 2/(e^{2t} - 1).$$

Thus

$$\frac{1}{G(t)} = \frac{1}{t}\left(t - t^2 + \frac{2t}{e^{2t} - 1}\right)$$

Similarly, with

$$H(t) = O(2)\frac{t^2}{2!} + O(4)\frac{t^4}{4!} + \cdots,$$

one obtains, from Theorem 1:

$$H'(t) - \big(t + t^2G(t)\big)H(t) = t + t^2G(t). \tag{2}$$

The solution of the first order linear differential equation (2) is

$$\begin{aligned} H(t) &= -1 + e^{t^2/2}\exp\left(\int_0^t \tau^2 G(\tau)\,d\tau\right) \\ &= -1 + \exp\left(\frac{t^2}{2} + \frac{O(1)}{1!}\frac{t^4}{4} + \frac{O(3)}{3!}\frac{t^6}{6} + \cdots\right). \quad \blacksquare \end{aligned}$$

REFERENCES

1. J. A. Bondy and U. S. R. Murty, Graph Theory with Applications, Macmillan, London, 1976.
2. R. L. Kimball, Graph Vertex Coloring Algorithms, Ph.D. Thesis, University of New Brunswick, 1984.
3. J. Riordan, An Introduction to Combinatorial Analysis, Wiley, New York, 1958.

Examples of Infinite, Incongruent Exact Covers

JOHN BEEBEE
Department of Mathematical Sciences, University of Alaska, Anchorage, AK 99508

Stein [1] asks: "If the set of integers is represented as the union of a denumerable number of (pairwise disjoint) incongruent arithmetic sequences $(d_i : b_i)$, $1 \leqslant i < \infty$, and $\Sigma d_i^{-1} = 1$, does it follow that $d_i = 2^i$?" I show by example that the answer is "no," and pose 3 similar problems.

The symbol $(d : b)$ represents the arithmetic sequence $\{x : x = b + ad,\ a \in Z\}$, where b is the initial position, d is the modulus of the sequence. A finite collection

of arithmetic sequences $\{(d_i : b_i)\}$ with the property that each integer belongs to exactly one sequence is called an exact cover. For every exact cover, $\Sigma d_i^{-1} = 1$.

A collection of arithmetic sequences is incongruent if no two moduli are the same. There are no finite incongruent exact covers (see [**3**]). A collection of arithmetic sequences satisfying the conditions in Stein's problem can be called an infinite incongruent exact cover, or IIEC. A collection that satisfies all the conditions except incongruence is called an IEC. A summary of results about exact covers and related systems is given in [**2**], which has an extensive bibliography.

It can easily be verified that the set of arithmetic sequences $\{(2^i : 2^{i-1}) : 1 \leqslant i < \infty\}$ is an IIEC, so Stein's problem asks if there is an IIEC with different moduli from this one. I call this particular IIEC the infinite Gray cover, because of its close connection with the binary Gray code.

DEFINITION. The Gray cover with parameters p and k has moduli

$$\underbrace{p, p, \ldots, p}_{\leftarrow k-1 \rightarrow}; \underbrace{p^2, p^2, \ldots, p^2}_{\leftarrow k-1 \rightarrow}; \ldots; \underbrace{p^k, p^k, \ldots p^k}_{\leftarrow k-1 \rightarrow}; p^k$$

and corresponding initial positions

$$1, 2, \ldots, p-1; p, 2p, \ldots, (p-1)p; \ldots; p^{k-1}, 2p^{k-1}, \ldots, (p-1)p^{k-1}; p^k.$$

The Gray cover with parameters p and k is an exact cover. If we let $p = 2$ and imagine k to be infinite, we get the previously mentioned infinite Gray cover. If we let $p = 3$ and imagine k to be infinite, then each integer belongs to precisely one of the arithmetic sequences

$$\{(3^i : 3^{i-1}) : 1 \leqslant i < \infty\} \cup \{(3^i : 2 \cdot 3^{i-1}) : 1 \leqslant i < \infty\}, \tag{1}$$

and

$$\sum_{i=1}^{\infty} \frac{2}{3^i} = 1,$$

so (1) is an IEC.

Construction. Let $C_1 = \{(d_i : b_i)\}, C_2 = \{(e_j : c_j)\}$ be exact covers. Then $C_3 = \{(d_i : b_i) : i \neq I\} \cup \{(e_j d_I : b_I + c_j d_I)\}$ is an exact cover. I call C_3 the exact cover formed by covering the Ith sequence of C_1 by C_2.

This construction can be extended somewhat. If C_1 or C_2 is an IEC, so is C_3. Also, the construction can be applied repeatedly and iteratively ($C_1 = C_2$).

Proof. To prove that C_3, constructed as described, really is an exact cover, let x be an integer. Since C_1 is an exact cover, there is a unique index i and number α such that $x = b_i + \alpha d_i$, so if $i \neq I$ then $x \in (d_i : b_i) \in C_3$. If $i = I$, then $x = b_I + \alpha d_I$, and C_2 is an exact cover, so there is a unique index j and number β such that $\alpha = c_j + \beta e_j$, and, hence,

$$x = b_I + (c_j + \beta e_j) d_I = b_I + c_j d_I + \beta(e_j d_I).$$

Thus, $x \in (e_j d_I : b_I + c_j d_I) \in C_3$. Since the indices i and j are unique for each x, x belongs to precisely one sequence in C_3. Thus C_3 is an exact cover.

If we cover each of the sequences $(3^i : 3^{i-1})$ in (1) by the infinite Gray cover, we get the collection

$$\{(3^i 2^j : 3^{i-1} + 2^{j-1} \cdot 3^i) : 1 \leqslant j < \infty, 1 \leqslant i < \infty\} \cup \{(3^i : 2 \cdot 3^{i-1}) : 1 \leqslant i < \infty\}. \quad (2)$$

The method of construction suggests that (2) is an IEC, and in fact it can be verified that it is an IIEC with moduli different from the infinite Gray cover. Obviously the smallest modulus is 3 and 3 is a divisor of all the moduli.

Using the same idea, we can construct an IIEC with smallest modulus 4, but 4 does not divide all the moduli. We begin with the infinite Gray cover $\{(2^i : 2^{i-1}) : 1 \leqslant i < \infty\}$ and cover the first sequence $(2 : 1)$ with the IIEC (2), getting the IEC

$$\begin{aligned} &\{(2 \cdot 3^i \cdot 2^j : 1 + 2 \cdot 3^{i-1} + 2^j 3^i) : 1 \leqslant j < \infty, 1 \leqslant i < \infty\} \\ &\quad \cup \{(2 \cdot 3^i : 1 + 4 \cdot 3^{i-1}) : 1 \leqslant i < \infty\} \cup \{(2^i : 2^{i-1}) : 2 \leqslant i < \infty\}, \end{aligned} \quad (3)$$

which is seen to be an IIEC.

Using the ideas outlined here, it is possible to construct an infinity of IIECS, all with smallest modulus 2, 3, or 4; and none with 4 a divisor of all the moduli.

I have not been able to solve the following problems.

1. Construct an IIEC with smallest modulus larger than 4. Note that if the smallest modulus is 5 or any prime, then it divides every modulus. In fact, it can be shown that if the smallest modulus is a prime p, then there are IIECs with smallest modulus q, for all $q < p$.
2. Construct an IIEC for which 1 is the greatest divisor of the moduli.
3. Let

$$\begin{aligned} d_{ij0} &= 2^i 3^j, & 1 &\leqslant i, j < \infty, \\ d_{0jk} &= 3^j 5^k, & 1 &\leqslant j, k < \infty, \\ d_{i0k} &= 2^i 5^k, & 1 &\leqslant i, k < \infty, \\ d_{ijk} &= 2^i 3^j 5^k, & 1 &\leqslant i, j, k < \infty. \end{aligned}$$

Then $\Sigma d_{ijk}^{-1} = 1$. Find b_{ijk} so that $(d_{ijk} : b_{ijk})$ is an IIEC, or show that none exist. A positive answer to this would solve 1 and 2.

REFERENCES

1. S. Stein, Unions of Arithmetic Sequences, *Math. Annalen*, 134 (1958) 282–294.
2. S. Porubsky, Results and Problems on Covering Systems of Residue Classes, Mitteilungen aus dem Mathem seminar Giessen, 1981.
3. B. Novák and S. Znám, Disjoint Covering Systems, *Amer. Math. Monthly*, 81 (1974) 42–45.

EDITED BY JOAN P. HUTCHINSON and STAN WAGON

The Absolute and Uniform Convergence of Infinite Improper Integrals

DAVID J. HALLENBECK AND KATARZYNA TKACZYŃSKA
Department of Mathematical Sciences, University of Delaware, Newark, DE 19716

In standard advanced calculus texts the most common test for establishing the uniform convergence of an infinite improper integral is the *Weierstrass M-test*: Suppose $f(x, t)$ is Riemann integrable over $[a, c]$ for all $c \geqslant a$ and all $x \in [\alpha, \beta]$. Suppose that there exists a positive function $M(t)$ defined for $t \geqslant a$ such that $|f(x, t)| \leqslant M(t)$ for $t \geqslant a$ and $x \in [\alpha, \beta]$, and such that $\int_a^\infty M(t)\,dt$ exists. Then for each $x \in [\alpha, \beta]$, $\int_a^\infty f(x, t)\,dt$ is absolutely convergent and the convergence is uniform on $[\alpha, \beta]$ [**1**, p. 268], [**3**, p. 348].

The Weierstrass M-test also gives the uniform convergence of $\int_a^\infty |f(x, t)|\,dt$. For this reason students often assume that if $\int_a^\infty f(x, t)\,dt$ converges uniformly and absolutely then $\int_a^\infty |f(x, t)|\,dt$ also converges uniformly. In this note we present an elementary example to show that this assumption is in general false. The following mean value property is the main tool in the proof of our theorem.

LEMMA. *Let f be a positive decreasing function on $[0, \infty)$ and let g be defined on $[0, \infty)$ by*

$$g(x) = \begin{cases} 1 & 2k \leqslant x < 2k+1 \qquad k = 0, 1, \dots \\ -1 & 2k-1 \leqslant x < 2k \qquad k = 1, 2, \dots \end{cases}.$$

Let $[a, b]$ be contained in $[0, \infty)$. Then there exists a real number $c \in [a, b]$ such that

$$\int_a^b f(t) g(t)\,dt = f(a) \int_a^c g(t)\,dt. \tag{1}$$

Proof. Let

$$A = \frac{\int_a^b f(t) g(t)\,dt}{f(a)} \quad \text{and} \quad G(x) = \int_a^x g(t)\,dt.$$

Since g is integrable on $[a, b]$ we know that $G(x)$ is a continuous function of x on $[a, b]$ [**2**, p. 133]. Hence G has the intermediate value property and (1) will hold if we show that

$$\min_{x \in [a, b]} G(x) \leqslant A \leqslant \max_{x \in [a, b]} G(x). \tag{2}$$

We call the intervals $[2k, 2k+1)$ even $(k = 0, 1, \dots)$ and the intervals $[2k-1, 2k)$ odd $(k = 1, 2, \dots)$. If $a = b$ then (1) holds trivially so suppose $a \neq b$. Clearly either a belongs to one of the even intervals or one of the odd intervals.

We first consider the case when a belongs to an even interval. Let a^* be the smallest natural number larger than a, and let $d = a^* - a$. It follows from the definition of g and the fact that f is decreasing that

$$-(1-d)f(a^*) < \int_a^b f(t)g(t)\,dt < df(a). \tag{3}$$

Since $-f(a^*) > -f(a)$, (3) implies

$$(-1+d)f(a) < \int_a^b f(t)g(t)\,dt < df(a), \tag{4}$$

and so we have $-1 + d < A < d$. If G attains the values $-1 + d$ and d in the interval $[a, b]$ then (2) holds and the proof is complete.

We next consider the case when $\max_{x\in[a,b]}G(x) < d$ and (or) $\min_{x\in[a,b]}G(x) > -1 + d$. Let $b^* = a^* + 1$ and note that if $x \in [a, b]$ then $-1 + d \leqslant G(x) \leqslant d$. Furthermore, $b \geqslant b^*$ implies that G attains both $-1 + d$ and d. Therefore we need only consider the case when $b \in (a, b^*)$. Note that 0 is always one of the values attained by G.

Assume $\max_{x\in[a,b]}G(x) = D$ and $0 < D < d$. Then $b < a^*$ since if $b \geqslant a^*$ the maximum value of G is d. Hence $b - a = D$ and this implies that $\int_a^b f(t)g(t)\,dt < Df(a)$. So we have $A < D$.

Now assume $\min_{x\in[a,b]}G(x) = \tilde{D}$ where $-1 + d < \tilde{D} < 0$. Then $b \in (a^*, b^*)$ and if we let $\tilde{a} = a^* + d$ we have $b - \tilde{a} = -\tilde{D}$. This implies that $\int_a^b f(t)g(t)\,dt > \tilde{D}f(a)$, and so $A > \tilde{D}$.

It follows that if $-1 + d$ is attained by G and $D < d$ then $-1 + d < A < D$. Also, if d is attained by G and $\tilde{D} > -1 + d$, then $\tilde{D} < A < d$. Finally, if $D < d$ and $\tilde{D} > -1 + d$ we have $\tilde{D} < A < D$. Hence,

$$\min_{x\in[a,b]} G(x) \leqslant A \leqslant \max_{x\in[a,b]} G(x),$$

and the proof is complete in the case a belongs to an even interval.

When a belongs to an odd interval substitute $-g$ for g and apply the even case to obtain

$$\int_a^b \{f(t)\}\{-g(t)\}\,dt = f(a)\int_a^c \{-g(t)\}\,dt \tag{5}$$

for some $c \in [a, b]$. This is equivalent to (1) and completes the proof.

The following theorem gives an example of a function $f(x, t)$ defined for $t \geqslant 0$ for which $\int_0^\infty f(x, t)\,dt$ converges uniformly on $x \geqslant 0$, $\int_0^\infty |f(x, t)|\,dt$ converges on $x \geqslant 0$, but the convergence of this last integral is not uniform for $x \geqslant 0$. This cannot occur in the cases where the Weierstrass M-test applies to $\int_0^\infty f(x, t)\,dt$ since in these cases we have $\int_0^\infty |f(x, t)|\,dt$ converging uniformly.

THEOREM. *If $g(t)$ is defined as in the previous lemma, then $\int_0^\infty xg(t)/(x^2 + t^2)\,dt$ converges uniformly for $x \geqslant 0$ while $\int_0^\infty |xg(t)/(x^2 + t^2)|\,dt$ converges but not uniformly on $x \geqslant 0$.*

Proof. Let $0 < a < b$ and note that by the previous lemma we have

$$\int_a^b \frac{x}{x^2 + t^2} g(t)\, dt = \frac{x}{x^2 + a^2} \int_a^c g(t)\, dt \tag{6}$$

for some $c \in [a, b]$. Since $|\int_a^c g(t)\, dt| \leqslant 1$, (6) implies that

$$\left| \int_a^b \frac{x}{x^2 + t^2} g(t)\, dt \right| \leqslant \frac{x}{x^2 + a^2}. \tag{7}$$

Since $x/(x^2 + a^2) \leqslant 1/2a$ for all $x \geqslant 0$, it follows from (7) that given any $\varepsilon > 0$ there exists a real number Q such that if $a \geqslant Q$ then

$$\left| \int_a^b \frac{x}{x^2 + t^2} g(t)\, dt \right| < \varepsilon. \tag{8}$$

It follows from the Cauchy criterion for uniform convergence [**1**, p. 267] that $\int_0^\infty xg(t)/(x^2 + t^2)\, dt$ converges uniformly on $x \geqslant 0$. It is easy to verify that if $F(x) = \int_0^\infty x/(x^2 + t^2)\, dt$ then $F(0) = 0$ and $F(x) = \pi/2$ if $x > 0$. Since $F(x)$ is not continuous on $[0, \infty)$ it follows that the convergence of $\int_0^\infty |xg(t)/(x^2 + t^2)|\, dt$ is not uniform on $[0, \infty)$ [**1**, p. 270]. This completes the proof.

REFERENCES

1. Robert G. Bartle, The Elements of Real Analysis, second ed., Wiley, New York, 1976.
2. Walter Rudin, Principles of Mathematical Analysis, third ed., McGraw Hill, New York, 1976.
3. David V. Widder, Advanced Calculus, second ed., Prentice-Hall, Englewood Cliffs, New Jersey, 1961.

Special Integration Techniques for Trigonometric Integrals

ASHOK K. ARORA, SUDHIR K. GOEL, AND DENNIS M. RODRIGUEZ
Department of Applied Mathematical Sciences, University of Houston-Downtown
Houston, TX 77002

In this paper we present two basic properties of definite integrals. These properties are in common use and can be easily proved. We use these properties to evaluate certain complicated integrals some of which are normally evaluated using contour integration.

Property 1. If f is an integrable function over the closed interval $[0, a]$, then

$$\int_0^a f(x)\, dx = \int_0^{a/2} f(x)\, dx + \int_0^{a/2} f(a - x)\, dx.$$

The following two facts are immediate from property 1. (A) If $f(a - x) = f(x)$, then

$$\int_0^a f(x)\, dx = 2\int_0^{a/2} f(x)\, dx;$$

and (B) If $f(a-x)=-f(x)$, then

$$\int_0^a f(x)\,dx=0.$$

One should observe that a function satisfying the hypothesis of fact (A) is symmetric over the interval $[0, a]$ across the line $x=a/2$, and a function satisfying the hypothesis of fact (B) is symmetric over the interval $[0, a]$ through the point $(a/2, 0)$.

Property 2. If f is an integrable function over the closed interval $[0, a]$, then

$$\int_0^a f(x)\,dx=\int_0^a f(a-x)\,dx.$$

We illustrate the use of the above properties with several examples.

The first example we present appears in most complex analysis books (see [**3**], [**4**]). The integral in this example is generally evaluated using contour integration and the procedure is quite involved (see [**4**, pp. 113–114]).

Example 1.

$$\int_0^\infty \frac{\ln x}{1+x^2}\,dx=0.$$

Solution. If we let I denote the value of this integral, then using the substitution $u=1/x$ and simplifying using $\ln(1/u)=-\ln u$, we obtain

$$I=-\int_0^\infty \frac{\ln u}{1+u^2}\,du=-I,$$

whence $I=0$.

An alternative way to solve example 1 is to use the trig substitution $x=\tan\theta$ and then use property 2.

Example 2.

$$\int_0^\infty \frac{\ln x}{(1+x^2)^2}\,dx=-\frac{\pi}{4}$$

Solution. Let $x=\tan\theta$; then $dx=\sec^2\theta\,d\theta$. Thus

$$\begin{aligned} I&=\int_0^\infty \frac{\ln x}{(1+x^2)^2}\,dx=\int_0^{\pi/2}\frac{\ln(\tan\theta)}{\sec^4\theta}\sec^2\theta\,d\theta\\ &=\int_0^{\pi/2}\cos^2\theta\ln(\tan\theta)\,d\theta\\ &=\int_0^{\pi/2}\left(\frac{1+\cos 2\theta}{2}\right)\ln(\tan\theta)\,d\theta\\ &=\frac{1}{2}\int_0^{\pi/2}\ln(\tan\theta)\,d\theta+\frac{1}{2}\int_0^{\pi/2}\cos 2\theta\ln(\tan\theta)\,d\theta. \end{aligned}$$

Now, $\int_0^{\pi/2} \ln(\tan\theta)\,d\theta = \int_0^{\pi/2} \ln(\sin\theta)\,d\theta - \int_0^{\pi/2} \ln(\cos\theta)\,d\theta = 0$ by property 2. We evaluate the second integral by parts, letting $u = \ln(\tan\theta)$ and $dv = \cos 2\theta\,d\theta$. Then

$$du = \frac{\sec^2\theta}{\tan\theta}\,d\theta = \frac{2}{\sin 2\theta}\,d\theta \quad \text{and} \quad v = \frac{1}{2}\sin 2\theta.$$

Therefore,

$$I = \left(\frac{1}{4}\sin 2\theta \ln(\tan\theta)\right)\Bigg|_0^{\pi/2} - \frac{1}{4}\int_0^{\pi/2} \sin 2\theta \cdot \frac{2}{\sin 2\theta}\,d\theta.$$

Using L'Hôpital's rule, we obtain $I = 0 - \pi/4 = -\pi/4$ as desired.

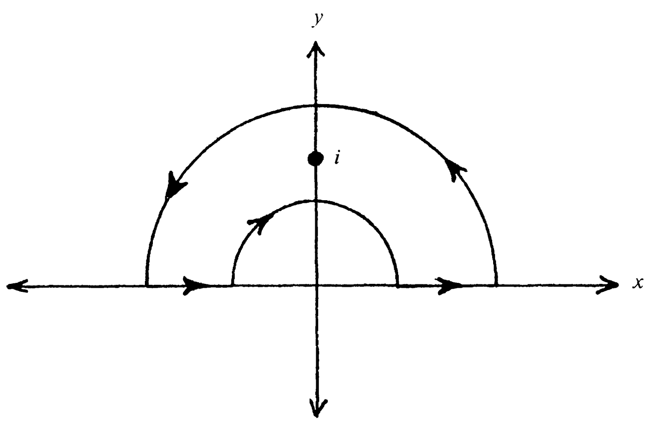
Examples 1 and 2 above appear as problems in [**3**, p. 183]. The hint provided to solve these problems is to use the branch

$$\ln z = \ln r + i\theta \qquad \left(r > 0, -\frac{\pi}{2} < \theta < \frac{3\pi}{2}\right)$$

of the logarithmic function, and integrate the functions

$$\frac{\ln z}{z^2+1} \text{ and } \frac{\ln z}{(z^2+1)^2}$$

around the closed contour shown below:



Our next example appears in the book *Complex Analysis* by Ahlfors [**1**, p. 159]. Its evaluation using contour integration is fairly involved. We noticed that this integral is important since it is the main result needed in proving Jensen's formula (see [**1**, p. 205]). The solution given in example 3 below also appears in [**5**].

Example 3.

$$\int_0^{\pi} \ln(\sin\theta)\,d\theta = -\pi \ln 2.$$

Solution. Using (A) above, we get

$$\int_0^{\pi} \ln(\sin\theta)\, d\theta = 2\int_0^{\pi/2} \ln(\sin\theta)\, d\theta.$$

Using property 2 above, we obtain

$$I = \int_0^{\pi/2} \ln(\sin\theta)\, d\theta = \int_0^{\pi/2} \ln(\cos\theta)\, d\theta.$$

Therefore,

$$\begin{aligned} 2I &= \int_0^{\pi/2} [\ln(\sin\theta) + \ln(\cos\theta)]\, d\theta \\ &= \int_0^{\pi/2} \ln\left(\frac{1}{2}\sin 2\theta\right) d\theta \\ &= \int_0^{\pi/2} \ln(\sin 2\theta)\, d\theta - \int_0^{\pi/2} \ln 2\, d\theta. \end{aligned}$$

Substituting $t = 2\theta$ in the first integral yields

$$2I = \frac{1}{2}\int_0^{\pi} \ln(\sin t)\, dt - \frac{\pi}{2}\ln 2.$$

Using (A) again, we have

$$2I = \int_0^{\pi/2} \ln(\sin t)\, dt - \frac{\pi}{2}\ln 2 = I - \frac{\pi}{2}\ln 2.$$

Therefore, $I = -(\pi/2)\ln 2$. Hence,

$$\int_0^{\pi} \ln(\sin\theta)\, d\theta = 2I = -\pi \ln 2,$$

as desired.

Our next example was proposed as a problem by Suresh Ailawadi [**2**, p. 439]. A solution, different from the one we present below, appeared in the September 1985 issue of the *College Mathematics Journal.*

Example 4. Evaluate

$$I = \int_0^{\pi/2} \frac{\sin^n x}{\sin^n x + \cos^n x}\, dx.$$

Solution. Using property 2 above, we get

$$I = \int_0^{\pi/2} \frac{\sin^n x}{\sin^n x + \cos^n x}\, dx = \int_0^{\pi/2} \frac{\cos^n x}{\cos^n x + \sin^n x}\, dx.$$

It follows that

$$2I = \int_0^{\pi/2} dx = \frac{\pi}{2}, \text{ whence } I = \frac{\pi}{4}.$$

Note that I is independent of n.

There exist many other integrals that can be evaluated using properties 1 and 2, three of which are listed below:

$$\int_0^{\pi} \ln(1 + \cos\theta)\, d\theta,$$

$$\int_0^{\infty} \ln\left(x + \frac{1}{x}\right) \frac{dx}{1 + x^2},$$

$$\int_0^{\pi} \frac{a^n \sin^2 x + b^n \cos^2 x}{a^{2n} \sin^2 x + b^{2n} \cos^2 x}\, dx.$$

REFERENCES

1. Lars V. Ahlfors, Complex Analysis, 2nd ed., McGraw-Hill, New York, 1966.
2. Suresh Ailawadi, Problem 260, *Two-Year College Mathematics Journal*, 14 (1983) 439.
3. Ruel V. Churchill and James W. Brown, Complex Variables and Applications, 4th ed., McGraw-Hill, New York, 1984.
4. John B. Conway, Functions of One Complex Variable, Springer-Verlag, New York, 1973.
5. Shanti Narayan, Integral Calculus, 9th ed., S. Chand and Co., Ram Nagar, New Delhi, 1975.

PROBLEMS AND SOLUTIONS

*For instructions about submitting **proposed** problems for publication in this department see the inside front cover. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the address given on the inside front cover. Two copies suffice.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

For instructions about submitting solutions of Problems, which should be mailed before June 30, 1988, see the inside front cover. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label if an acknowledgment is desired.

E 3249. *Proposed by Toma Albu, University of Bucharest, Romania.*

Show that, for all primes p and all natural numbers m, the sum

$$\sum_{k=1}^{m} \frac{k}{1+k(p-1)}$$

is not an integer.

E 3250. *Proposed by Lennart Bondesson, Umea, Sweden.*

Suppose $N > 1$ and $a_1 < a_2 < \cdots < a_N$ are given real numbers. Consider the partial fraction decomposition

$$\prod_{i=1}^{N} \frac{1}{a_i + x} = \sum_{i=1}^{N} \frac{A_i}{a_i + x}.$$

It is familiar that the numbers $A_1, A_2, \ldots$ alternate in sign. What can we say about the signs of the partial sums $A_1, A_1 + A_2, A_1 + A_2 + A_3, \ldots, A_1 + A_2 + \cdots + A_N$?

E 3251. *Proposed by Clark Kimberling, University of Evansville, Indiana.*

Given a triangle ABC, it is possible to construct three circles Γ_A, Γ_B, Γ_C inside ABC in such a way that each one is externally tangent to the other two, Γ_A is

tangent to the two sides AB and AC, Γ_B is tangent to the two sides BC and BA, and Γ_C is tangent to the two sides CA and CB. (Cf. A. S. Winsor, *Modern Higher Plane Geometry*, Christopher Publishing House, Boston, 1941, pp. 206–209 or H. G. Forder, *Higher Course in Geometry*, pp. 244–245.) Let D be the point of contact of Γ_B and Γ_C, E the point of contact of Γ_C and Γ_A, and F the point of contact of Γ_A and Γ_B.

(a) Prove that the lines AD, BE, and CF are concurrent.

*(b) Prove that the lines $A'D$, $B'E$, and $C'F$ are concurrent, where A', B', and C' are the excenters of the triangle ABC.

E 3252. *Proposed by David G. Winslow, Louisiana State University, Baton Rouge, Louisiana.*

(a) Suppose that $h\colon R \to R$ is continuous and that h^n is bounded for some positive integer n, where $h^2(x) = h(h(x))$, $h^3(x) = h(h(h(x)))$, etc. Prove that either h or h^2 is bounded.

(b) Show that for each positive integer n there is a continuous function $h\colon R^2 \to R^2$ such that h^k is unbounded for $1 \leqslant k \leqslant n-1$ but h^n is bounded.

SOLUTIONS OF ELEMENTARY PROBLEMS

Brocard Revisited

E 2905* [1981, 619]. *Proposed by R. J. Stroeker, Erasmus Universiteit, Netherlands.*

Inside any triangle with vertices A, B, and C, a point P exists such that $\sphericalangle PAB = \sphericalangle PBC = \sphericalangle PCA =: \omega$. The point P is called a Brocard point and the angle ω is called its Brocard angle. If α, β, and γ are the angles of triangle ABC, then prove the inequalities:

$$\omega^{-1} < \alpha^{-1} + \beta^{-1} + \gamma^{-1} \leqslant \frac{3}{2}\omega^{-1} \quad \text{and} \quad \frac{3}{4}\omega^{-2} \leqslant \alpha^{-2} + \beta^{-2} + \gamma^{-2} < \omega^{-2}.$$

Editorial Comment. Partial solutions were received from P. S. Bruckman, H. H. Guggenheimer, and the proposer. Later the following relevant paper appeared:

R. J. Stroeker and H. J. T. Hoogland, "Brocardian geometry revisited or some remarkable inequalities," *Nieuw Archief voor Wiskunde* (4) 2(1984), 281–310.

This paper gives complete proofs for the two assertions:

$$\alpha^{-1} + \beta^{-1} + \gamma^{-1} > \omega^{-1}, \qquad \alpha^{-2} + \beta^{-2} + \gamma^{-2} < \omega^{-2}$$

and provides numerical evidence in support of the other two inequalities

$$\alpha^{-1} + \beta^{-1} + \gamma^{-1} \leqslant \frac{3}{2}\omega^{-1}, \qquad \alpha^{-2} + \beta^{-2} + \gamma^{-2} \geqslant \frac{3}{4}\omega^{-2}.$$

These last two conjectures were proved by V. Mascioni in the following two papers:

Zur Abschätzung des Brocardschen Winkels, Elem. Math. 41 (1986) 98–101;

Zur Abschätzung des Brocardschen Winkels II, Elem. Math. 42 (1987) 35–42.

Improved versions of the inequalities of the problem have been obtained in a forthcoming paper of F. F. Abi-Khuzam and A. B. Boghossian. The relevant parts of their results may be summarized as follows:

$$\omega^{-1} + 3\pi^{-1} \leqslant \alpha^{-1} + \beta^{-1} + \gamma^{-1} \leqslant \frac{3}{2}\omega^{-1}; \tag{1}$$

$$\frac{3}{4}\omega^{-2} \leqslant \alpha^{-2} + \beta^{-2} + \gamma^{-2} \leqslant \omega^{-2} - 9\pi^{-2}; \tag{2}$$

equality holds in each case if and only if the triangle is equilateral. They also show that there exists a number $\delta \in (1, 2)$ such that

$$\alpha^{-\lambda} + \beta^{-\lambda} + \gamma^{-\lambda} \leqslant 3(2\omega)^{-\lambda} \quad \text{for all} \quad \lambda \in [0, \delta]; \quad \text{and}$$
$$\alpha^{-\lambda} + \beta^{-\lambda} + \gamma^{-\lambda} \geqslant 3(2\omega)^{-\lambda} \quad \text{for all} \quad \lambda \notin (0, \delta).$$

Sums of Consecutive Square Roots

E 3010 [1983, 483]. *Proposed by F. David Hammer, Redwood City, CA.*

Prove that for all $n \geqslant 0$ the relation

$$\left[n^{1/2} + (n + 1)^{1/2} + (n + 2)^{1/2}\right] = \left[(9n + 8)^{1/2}\right]$$

holds. Here $[\cdot]$ denotes the greatest integer function.

Solution I by J. C. Binz, University of Bern, Switzerland. Let $x = \sqrt{n} + \sqrt{n+1} + \sqrt{n+2}$. Then

$$x^2 = 3n + 3 + 2\left(\sqrt{n(n+1)} + \sqrt{n(n+2)} + \sqrt{(n+1)(n+2)}\right).$$

For $n \geqslant 1$ the inequalities $(n + 2/5)^2 < n(n + 1) < (n + 1/2)^2$, $(n + 7/10)^2 < n(n + 2) < (n + 1)^2$, and $(n + 7/5)^2 < (n + 1)(n + 2) < (n + 3/2)^2$ lead to $9n + 8 < x^2 < 9n + 9$, and, therefore, $[x] = [\sqrt{9n + 8}]$; the case $n = 0$ is verified directly.

Solution II by James Propp, Berkeley, California. The relation is easily checked by hand in the cases $n = 0$, 1, and 2. When $n \geqslant 3$, we have

$$n(n + 1)(n + 2) > \left(n + \frac{8}{9}\right)^3$$

(straightforwardly verifiable by estimating the roots of the quadratic polynomial $(x - 1)x(x + 1) - (x - 1/9)^3$), so that by the arithmetic mean-geometric mean

inequality,

$$\frac{\sqrt{n}+\sqrt{n+1}+\sqrt{n+2}}{3} > \sqrt[3]{\sqrt{n}\sqrt{n+1}\sqrt{n+2}} > \sqrt{n+\frac{8}{9}}$$

and

$$\sqrt{n}+\sqrt{n+1}+\sqrt{n+2} > \sqrt{9n+8}\,.$$

But by the root-mean-square inequality,

$$\frac{\sqrt{n}+\sqrt{n+1}+\sqrt{n+2}}{3} < \sqrt{\frac{n+(n+1)+(n+2)}{3}}$$

and

$$\sqrt{n}+\sqrt{n+1}+\sqrt{n+2} < \sqrt{9n+9}\,.$$

Since no integer can lie between $\sqrt{9n+8}$ and $\sqrt{9n+9}$, both $\sqrt{n}+\sqrt{n+1}+\sqrt{n+2}$ and $\sqrt{9n+8}$ must have the same integer part.

A solution involving the application of Taylor series methods (for $\sqrt{n+x}$) was submitted by O. P. Lossers. A. M. Rockett showed that both $\sqrt{9n+8}$ and $\sqrt{n}+\sqrt{n+1}+\sqrt{n+2}$ (but no integer) lie between $\sqrt{9n+7}$ and $\sqrt{9n+9}$. A. Tissier (France) showed that if $a_n=\sum_{q=-k}^{k}(n+q)^{1/2}$ and $b_n=\{(2k+1)^2n-1\}^{1/2}$, then $[a_n]=[b_n]$ for all sufficiently large n. In the last assertion P. D. McCray showed that the powers $1/2$ and 2 could be replaced with r and $1/r$, respectively, and R. B. Nelsen showed that $n+q$ could be replaced with $n+hq$ where h is an integer.

Also solved by 39 other readers and the proposer.

Ups and Downs

E 3042 [1984, 257]. *Proposed by A. M. Fink, Iowa State University, and Max Jodeit, Jr., University of Minnesota.*

Which real-valued functions h defined on $[0,1]$ can be expressed as $h(x)=f(x)+g(x)$, where $f(x)\geqslant 0$ is nonincreasing, and $g(x)\geqslant 0$ is nondecreasing?

Solution by the proposers. A function h can be so expressed if and only if $V(h)\leqslant h(0)+h(1)$, where $V(h)$ is the variation of h on $[0,1]$.

If the condition holds, the decomposition $h(x)=(h(0)-N(x))+P(x)$ works ($P(x)$ and $N(x)$ are, respectively, the positive and negative variations of h on $[0,x]$). For $P(1)+N(1)=V(h)\leqslant h(0)+h(1)=2h(0)+P(1)-N(1)$, which implies $N(x)\leqslant N(1)\leqslant h(0)$. This decomposition is the only one with $g(0)=0$,

$f(1) = (1/2)(h(0) + h(1) - V(h))$ (by the minimum property of the Jordan decomposition).

If $h = f + g$, as in the problem, then

$$\begin{aligned} V(h) &\leqslant V(f) + V(g) = f(0) - f(1) + g(1) - g(0) \\ &= h(0) + h(1) - 2f(1) - 2g(0) \leqslant h(0) + h(1). \end{aligned}$$

Thus, if $V(h) = h(0) + h(1)$, $g(0) = 0 = f(1)$, so f and g are unique in this case.

These functions may oscillate; $(1/2)(1 + \cos^2 2\pi x)$ is an example. It is always possible to add cx to $g(x)$, $c \geqslant 0$, and subtract cx from $f(x)$ to get $h(x) = f_0(x) + g_0(x)$ with $g_0(0) = 0 = f_0(1)$.

Also solved by G. Behrendt (West Germany), A. Bondesen (Denmark), R. Davies, T. Jager, B. G. Klein, M. Meyerson, W. A. Newcomb, E. Ordman, R. Pemantle, S. Philipp, R. Shepler and V. Konecny, J. C. Tripp, U. of South Alabama Problem Group, and P. Y. Wu (Republic of China).

A Minimum Length by Fours

E 3049 [1984, 437]. *Proposed by Jordi Dou, Barcelona, Spain.*

Determine a planar region of area 4 which can be partitioned into four subregions of unit area in such a way that the total length of all bounding arcs is a minimum. For example, a square of area 4 can be partitioned into four unit squares so that the bounding arcs have total length 12, while a circle of area 4 can be partitioned into four sectors of unit area so that the total length of the bounding arcs is

$$(4\pi + 8)/\sqrt{\pi} \doteq 11.6.$$

Solution. The only solution that appears correct was that of the proposer. He asserts that the minimizing figure must satisfy the following conditions.

(i) All the boundary arcs in the minimizing figure must be straight segments or arcs of circles.

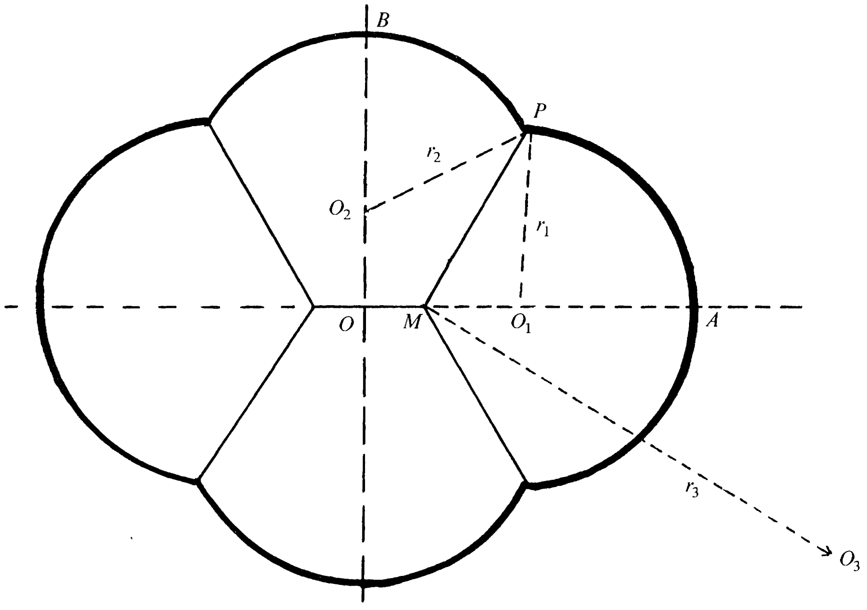
(ii) At each vertex there meet three arcs forming equal (120°) angles.

(iii) The appropriately signed curvatures of the three arcs meeting at a vertex sum to zero.

(iv) The figure has two perpendicular axes of symmetry.

The first three conditions are shown to be necessarily satisfied by an optimal figure in the paper "Isoperimetric problem with application to the figure of cells" by H. P. McKean, M. Schreiber, and G. H. Weiss, *Journal of Mathematical Physics*, 6(1965) 479–484. The fourth condition was not proved, but was supported by some heuristic reasons.

Assuming that the minimizing figure is as in the diagram below, its proportions are calculated as follows.



Let O_1 and O_2 be the centers of $\widehat{PA}$ and $\widehat{PB}$ and r_1, r_2 the respective radii. Let r_3 be the radius of $\widehat{MP}$ and 2ε (in degrees) the central angle subtended at center O_3. Assume for the while that chord $\overline{MP}$ has length 1; the figure will be renormalized below.

Then $r_3 = 1/(2\sin\varepsilon)$, $r_1 = \sin(60-\varepsilon)/\sin(90-2\varepsilon)$. From (iii) we have $1/r_3 = 1/r_1 - 1/r_2$, so $r_2 = r_1/(1-2r_1\sin\varepsilon)$.

Let α be the area of the region $MAPM$, so $\alpha = S_1 + T_1 + \sigma$, where S_1 is the area of the sector AO_1P, T_1 the area of triangle O_1PM, and σ the area of the segment bounded by $\widehat{MP}$ and $\overline{MP}$. Let β be the area of region $OMPBO = S_2 + T_2 + T - \sigma$, where S_2 is the area of sector BO_2P, T_2 that of triangle PO_2M, and T that of triangle O_2OM.

Then $S_1 = \pi r_1^2(45-\varepsilon)/180$, $T_1 = \sin(30-\varepsilon)r_1/2$, $\sigma = (\pi\varepsilon - 90\sin 2\varepsilon)r_3^2/180$, $S_2 = \pi r_2^2(30+\varepsilon)/180$, $T_2 = \sin(30+\varepsilon)r_2/2$, $T = [\sin(60-\varepsilon) - r_2\sin(30-2\varepsilon)]\cdot[r_2\sin(60+2\varepsilon) - \sin(30+\varepsilon)]/2$. The condition $\alpha(\varepsilon) = \beta(\varepsilon)$ implies $\varepsilon \doteq 1^\circ\!.92349794$. The total area $S = 8\alpha = 8\beta \doteq 5.99867234$. The total length

$$\begin{aligned} L &= 4\left(\widehat{AP} + \widehat{BP} + \widehat{MP}\right) + 2\overline{OM} \\ &= 4\left[\pi r_1(45-\varepsilon)/90 + \pi r_2(30+\varepsilon)/90 + \pi r_3\varepsilon/90\right] \\ &\quad + 2\left[r_2\sin(60+2\varepsilon) - \sin(30+\varepsilon)\right]. \end{aligned}$$

$L \doteq 13.70068728$. The total length λ for a total area $\Sigma = 4$ is $\lambda = \sqrt{4/S}\cdot L = K\cdot L \doteq 11.1878022$.

The corresponding parts of the rescaled figure (with the same lettering) are $K = \overline{MP} \doteq 0.816586932$; $\overline{OM} \doteq 0.22949299$; $\overline{MO_1} \doteq 0.385194625$; $OO_2 \doteq 0.36835908$; $r_1 \doteq 0.694647354$; $r_2 \doteq 0.736718161$; $r_3 \doteq 12.1642382$.

($\alpha(\varepsilon) = \beta(\varepsilon)$ was solved with a pocket Casio FX-702P.)

Three nonoptimal configurations were also received.

Where *I* Is

E 3071 [1985, 58]. *Proposed by K. Satyanarayana, Hyderabad, India.*

If in triangle ABC, angle $C >$ angle $B >$ angle A, then prove that I lies inside triangle OBH where, as usual, O, I, H denote the circumcenter, incenter and orthocenter, respectively, of triangle ABC.

Solution by O. P. Lossers, Eindhoven University of Technology, The Netherlands. Let α, β, and γ denote the angles of the triangle. Because $AH \perp BC$ and $BH \perp AC$, we find

$$\angle CAH = \angle CBH = \left|\frac{1}{2}\pi - \gamma\right|.$$

Furthermore, because ABO is isosceles,

$$\angle BAO = \angle ABO = \left|\frac{1}{2}(\pi - 2\gamma)\right| = \left|\frac{1}{2}\pi - \gamma\right|.$$

Therefore, AI is simultaneously the angular bisector of $\angle CAB$ and of $\angle HAO$. In the same way, BI is the angular bisector of $\angle CBA$ and $\angle HBO$. Let AI intersect OH in P and let BI intersect OH in Q. Evidently $OA = OB$, and $AH > BH$ because $\alpha < \beta$. Now we apply the well-known partition theorem of angular bisectors

$$AH : AO = PH : PO, \qquad BH : BO = QH : QO.$$

So we see that $PH > QH$, or that O, P, Q, and H are in that order on OH, i.e., I lies inside triangle OBH.

Also solved by L. Bankoff, J. Dou (Spain), N. Elkies, J. Fukuta (japan), W. Janous (Austria), L. Kuipers (Switzerland), W. A. Newcomb, C. R. Pranesachar (India), M. Vowe (Switzeland), and the proposer.

It was brought to the attention of the editors that this problem appeared as Problem 794 in Math. Mag., where it was proposed by L. Bankoff. Also see the article by A. P. Guinand, "Euler Lines, Tritangent Centers, and Their Triangles", this MONTHLY, 91 (1984) 290-300.

An Integral Root Test

E 3072 [1985, 58]. *Proposed by Louis Funar, Craiova, Romania.*

Let $P(x) = a_0 + a_1x + \cdots + a_nx^n$ be a real polynomial of degree $n \geqslant 2$ such that

$$0 < a_0 < -\sum_{k=1}^{[n/2]} \frac{1}{2k+1} a_{2k}.$$

Prove that $P(x)$ has a real zero r such that $|r| < 1$.

Solution by K. F. Andersen, University of Alberta, Edmonton, Canada. Since

$$\int_{-1}^{1} P(x)\,dx = \sum_{k=0}^{[n/2]} 2a_{2k}/(2k+1)$$

the hypothesis shows that the integral of $P(x)$ over the interval $(-1,1)$ is negative. It follows that $P(x) < 0$ for some $x \in (-1,1)$. Since $P(0) = a_0 > 0$, the Intermediate Value Theorem shows that $P(x)$ has a real zero in $(-1,1)$.

Also solved by 27 other readers.

Ptolemy Applied

E 3073 [1985, 147]. *Proposed by Calin P. Popescu, Bucharest, Romania.*

In the hexagon $A_1A_2A_3A_4A_5A_6$ the triangles $A_1A_3A_5$ and $A_2A_4A_6$ are equilateral. Is it true that $A_kA_{k+3}(k = 1,2,3)$ are congruent if their sum equals the perimeter of the hexagon? Also, consider the converse.

Solution by Jiro Fukuta, Shinsei, Japan. In quadrilateral $A_1A_2A_3A_5$, the following inequality holds:

$$A_1A_2 \cdot A_3A_5 + A_2A_3 \cdot A_5A_1 \geqslant A_1A_3 \cdot A_2A_5,$$

and so we have

$$A_1A_2 + A_2A_3 \geqslant A_2A_5. \tag{1}$$

Similarly,

$$A_3A_4 + A_4A_5 \geqslant A_4A_1. \tag{2}$$

$$A_5A_6 + A_6A_1 \geqslant A_6A_2. \tag{3}$$

Using (1), (2), and (3), we obtain

$$(\text{perimeter of the hexagon}) \geqslant A_1A_4 + A_2A_5 + A_3A_6. \tag{4}$$

If equality holds in (4), then equality also holds in (1), (2), and (3). If equality holds in (1), A_2 is on the circumcircle of the triangle $A_1A_3A_5$. Similarly, A_4 and A_6 are on the same circle. Thus, it is easy to prove that $A_kA_{k+3}(k = 1,2,3)$ are congruent.

Also solved by the proposer.

A Symmetric Inequality in Three Variables

E 3074 [1985, 147]. *Proposed by M. J. Pelling, London, England.*

Prove the inequality

$$\sum (1+p^2)^2(1+q^2)^2(p-r)^2(q-r)^2 \geqslant 2\prod(1+p^2)\prod(p-q)^2, \qquad p,q,r \in \mathbb{R}.$$

where the sums and products are over cyclic permutations of p, q, r and determine all the cases where equality holds.

Composite solution by Jiro Fukuta, Shinsei, Japan, and the proposer. Define a, b, c as follows:

$$\begin{aligned} a &= (1+p^2)(1+q^2)(r-p)(q-r) \\ b &= (1+q^2)(1+r^2)(p-q)(r-p) \\ c &= (1+r^2)(1+p^2)(q-r)(p-q). \end{aligned}$$

Substituting into $(a+b+c)^2 = \Sigma a^2 + 2\Sigma ab \geqslant 0$ we get

$$\begin{aligned} &\left\{\sum(1+p^2)(1+q^2)(r-p)(q-r)\right\}^2 \\ &\quad = \sum(1+p^2)^2(1+q^2)^2(r-p)^2(q-r)^2 \\ &\qquad + 2\left\{\prod(1+p^2)\right\}\left\{\prod(p-q)\right\}\left\{\sum(1+q^2)(r-p)\right\} \\ &\quad = \sum(1+p^2)^2(1+q^2)^2(r-p)^2(q-r)^2 - 2\prod(1+p^2)\prod(p-q)^2 \geqslant 0 \end{aligned}$$

which is the required inequality.

Equality holds if and only if $a+b+c=0$, i.e.,

$$\sum(1+p^2)(1+q^2)(r-p)(q-r) = 0. \tag{1}$$

Introducing parameters $x = (rq+1)/(q-r)$, $y = (rp+1)/(r-p)$, we have

$$r-p = \frac{r^2+1}{y+r}, \qquad q-r = \frac{r^2+1}{x-r}, \qquad p-q = \frac{-(r^2+1)(x+y)}{(y+r)(x-r)},$$

$$1+p^2 = \frac{(r^2+1)(1+y^2)}{(y+r)^2}, \qquad 1+q^2 = \frac{(r^2+1)(1+x^2)}{(x-r)^2},$$

which, on substituting into (1) and simplifying, gives

$$\frac{(r^2+1)^4}{(y+r)^3(x-r)^3}\left\{\left[1-xy(x^2+y^2+xy+2)\right] - r(x+y)^2(x-y)\right\} = 0, \tag{2}$$

so that the general case of equality in parametric form is given by

$$r = \frac{1-xy(x^2+y^2+xy+2)}{(x-y)(x+y)^2}, \qquad p = \frac{ry-1}{y+r}, \qquad q = \frac{rx+1}{x-r} \tag{3}$$

where in (3) $x \neq \pm y$. There is also a special case in (2) when $x = y = \pm 1/\sqrt{3}$ and r is arbitrary, except $r \neq \pm 1/\sqrt{3}$, with

$$p = \frac{r-\sqrt{3}}{1+r\sqrt{3}}, \qquad q = \frac{r+\sqrt{3}}{1-r\sqrt{3}} \quad \text{or} \quad p = \frac{r+\sqrt{3}}{1-r\sqrt{3}}, \qquad q = \frac{r-\sqrt{3}}{1+r\sqrt{3}}. \tag{4}$$

Finally, there is the case $p = q = r$, not included above because x, y are assumed finite.

The inequality was originally discovered in trigonometric form as

$$\sum \cos^2 \alpha \operatorname{cosec}^2(\beta - \gamma) \geqslant 2 \quad (\text{sum cyclically over } \alpha, \beta, \gamma),$$

which reduces to the algebraic version of the problem on setting $p = \tan \alpha$, $q = \tan \beta$, $r = \tan \gamma$ and clearing fractions. The parameters x, y of (3) are $x = \cot(\beta - \gamma)$ and $y = \cot(\gamma - \alpha)$.

Continuous Functions are Constant

E 3077 [1985, 148]. *Proposed by Nicholas Passell, University of Wisconsin-Eau Claire.*

Let $\mathbb{N}$ be the natural numbers with the topology T consisting of $\varnothing$, $\mathbb{N}$ and the complements of finite sets. Characterize the continuous functions from $\mathbb{R}$, the reals with the usual topology, to $\mathbb{N}$ with the topology T.

Solution by Alan Schuchat, Wellesley College. A function f from $\mathbb{R}$ with the usual topology to $\mathbb{N}$ with T is continuous if and only if f is constant.

Proof. Constant functions are always continuous. Conversely, if f is continuous then $\mathbb{R}$ is the disjoint union of the closed sets $F_j = f^{-1}(\{j\})$, for j in $\mathbb{N}$. By a connectedness and category argument (e.g., F. Hausdorff, *Set Theory*, Chelsea, 1957, p. 178), only one set F_j is nonempty. So f is constant.

Also solved by M. Bowron, F. S. Cater, Chico Problem Group, J. A. Frohliger and A. L. Thorsen, J. W. Grossman & D. Schmidt, D. Hamlin (student), G. M. Kellar, J.-C. Leccia (France), N. J. Lord (England), O. P. Lossers (The Netherlands), L. A. Lucas, M. D. Meyerson, W. A Newcomb, M. J. Reed & C. R. Diminnie, C. R. Rosentrater, T. Salat (Czechoslovakia), F. Siwiec, T. A. Summers, N. Y. Wong (Hong Kong), and the proposer.

Comment. Cater, Frohliger and Thorsen, and Lucas pointed out that $\mathbb{N}$ can be replaced by any countable T_1-space.

An Application of Gauss-Lucas

E 3078 [1985, 215]. *Proposed by Zalman Rubinstein, University of Colorado, Boulder, and the University of Haifa.*

What is the largest positive integer n for which the nonzero solutions of the equation $(1 + z)^n = 1 + z^n$ all lie on the unit circle?

Solution by W. O. Egerland, University of Baltimore. Let $p_{n-1}(z) = (z + 1)^n - z^n - 1$. The derivative $p'_{n-1}(z) = n((z + 1)^{n-1} - z^{n-1})$ has a root at $z = (\exp(2\pi i/(n - 1)) - 1)^{-1}$. Since, if $n \geqslant 8$, $|z| = (2\sin\{\pi/(n - 1)\})^{-1} > 1$, $p_{n-1}(z)$ cannot have all nonzero roots on the unit circle by the theorem of Gauss-Lucas; and

since $p_6(z) = 7z(z+1)(z^2+z+1)^2$, $n = 7$ is the largest positive integer for which all nonzero roots of $p_{n-1}(z)$ lie on the unit circle.

Also solved by C. K. Caldwell, B. P. Carter, N. Elkies, Z. Franco (student), D. Hamlin (student), A. A. Jagers (Netherlands), W. Janous (Austria), S. V. Kanetkar, K. Kearnes, I. E. Leonard, N. J. Lord (England), O. P. Lossers (Netherlands), J. M. Monier (France), K. L. Stellmacher, G. Sylvester, and the proposer.

A Congruence Argument

E 3080 [1985, 215]. *Proposed by Loren C. Larson, St. Olaf College.*

Can the following equations be satisfied with integers?

$$(x+1)^2 + a^2 = (x+2)^2 + b^2 = (x+3)^2 + c^2 = (x+4)^2 + d^2.$$

Solution by Mark Bowron, Charlottesville, Virginia. The equations cannot be satisfied with integers. For any integer h we have

$$h^2 \equiv \begin{cases} 0 \pmod 8 & \text{if } h \equiv 0 \pmod 4, \\ 1 \pmod 8 & \text{if } h \equiv 1 \text{ or } 3 \pmod 4, \\ 4 \pmod 8 & \text{if } h \equiv 2 \pmod 4. \end{cases}$$

Hence,

$$h^2 + k^2 \equiv \begin{cases} 0, 1, \text{ or } 4 \pmod 8 & \text{if } h \equiv 0 \pmod 4, \\ 1, 2, \text{ or } 5 \pmod 8 & \text{if } h \equiv 1 \text{ or } 3 \pmod 4, \\ 0, 4, \text{ or } 5 \pmod 8 & \text{if } h \equiv 2 \pmod 4, \end{cases}$$

for any integers h and k. Therefore, since $x+1, x+2, x+3, x+4$ form a complete set of residues modulo 4, the congruences

$$(x+1)^2 + a^2 \equiv (x+2)^2 + b^2 \equiv (x+3)^2 + c^2 \equiv (x+4)^2 = d^2 \equiv n \pmod 8$$

are satisfied only if $n \in \{0,1,4\} \cap \{1,2,5\} \cap \{0,4,5\}$, which is impossible.

Also solved by J. Binz (Switzerland), R. Breusch, D. M. Broline, C. K. Caldwell, P.-C. Chuang, N. Elkies, Z. Franco (student), C. Groenewoud, J. H. Halton, R. Heller, J. Hook, I. M. Isaacs, F. H. Kierstead, Jr., B. G. Klein, L. Kuipers (Switzerland), O. P. Lossers (The Netherlands), S. Marivani, J.-M. Monier (France), D. Nuenschwander (student, W. Germany), M. Nutt (Canada), N. Plotkin, R. A. Simon Chile), U. of South Alabama Problem Group, J. T. Ward, K. Woerner, and the proposer.

An Odd / Even Power Sum Inequality

E 3102* [1985, 507]. *Proposed by J. P. Lambert, University of Alaska, Fairbanks.*

Prove or disprove: For all integers $m, n \geqslant 1$,

$$\frac{\sum_{j=1}^m (2j-1)^{2n}}{\sum_{j=1}^m (2j)^{2n}} < \left(\frac{2m}{2m+1}\right)^{2n+1}.$$

No correct solutions were received. The following argument was prepared by Harold G. Diamond of the editorial staff. (It is the policy of the MONTHLY Problem Section to provide solutions whenever possible. In the case of the present problem, this policy was severely tested.)

The stated inequality is true. Indeed we have

THEOREM 1. *Let*

$$\Delta_\nu(m) \overset{\text{def}}{=} \left(\frac{2m}{2m+1}\right)^{\nu+1} - \sum_{j=1}^{m}(2j-1)^\nu \Big/ \sum_{j=1}^{m}(2j)^\nu. \tag{1}$$

Then $\Delta_\nu(m) > 0$ *for all positive integers* m *and all real* $\nu \geqslant 4$. *Also*, $\Delta_2(m) > 0$, $\Delta_3(m) > 0$, *and* $\Delta_1(m) < 0$ *for all positive integers* m.

We shall later give an asymptotic estimate, due to Paul T. Bateman, for $\Delta_\nu(m)$ for fixed ν as $m \to \infty$. This will show that if m is sufficiently large, depending on ν, then $\Delta_\nu(m) < 0$ for each real $\nu < 2$ and $\Delta_\nu(m) > 0$ for each real $\nu \geqslant 2$. Possibly $\Delta_\nu(m) > 0$ holds for each $\nu \geqslant 2$ with no restriction on m.

The inequality of Theorem 1 is remarkably sharp, considering its simplicity and uniformity. The identities that follow (3) (below) show that $\Delta_\nu(m) = O(m^{-3})$ as $m \to \infty$ for $\nu = 2, 3$. The asymptotic analysis will show that this O estimate holds for each $\nu \geqslant 2$ with an O-constant that depends on ν.

We begin the proof of (1) by noting that for relatively large ν a simpler inequality gives a better estimate. We have

$$\frac{\sum_{j=1}^{m}(2j-1)^\nu}{\sum_{j=1}^{m}(2j)^\nu} \leqslant \left(\frac{2m-1}{2m}\right)^\nu < \left(\frac{2m}{2m+1}\right)^{\nu+1} \qquad (\nu \geqslant 2m). \tag{2}$$

The first inequality holds for all positive values of ν since

$$\frac{1}{2^\nu} < \frac{3^\nu}{4^\nu} < \cdots < \frac{(2m-1)^\nu}{(2m)^\nu}.$$

The second inequality holds for $\nu \geqslant 2m$ by the binomial expansion

$$\left(1 - \frac{1}{4m^2}\right)^{-\nu} > 1 + \frac{\nu}{4m^2} \geqslant 1 + \frac{1}{2m}.$$

We shall henceforth assume that $\nu < 2m$.

Inequality (1) is equivalent to showing

$$\left(\frac{2m}{2m+1}\right)^{\nu+1} + 1 - \frac{\sum_{j=1}^{2m} j^\nu}{\sum_{j=1}^{m}(2j)^\nu} > 0. \tag{3}$$

We verify the cases $\nu = 2$ and $\nu = 3$ directly: Familiar formulas for sums of squares

and sums of cubes show that the left-hand side of (3) is

$$\left(\frac{2m}{2m+1}\right)^3 + 1 - \frac{4m+1}{2m+2} = \frac{4m+1}{(2m+1)^3(2m+2)} \qquad (\nu = 2)$$

$$\left(\frac{2m}{2m+1}\right)^4 + 1 - \frac{(2m+1)^2}{2(m+1)^2} = \frac{16m^3+22m^2+8m+1}{2(m+1)^2(2m+1)^4} \qquad (\nu = 3).$$

Henceforth, we assume $\nu \geqslant 4$.

We proceed by induction on m. In view of (2), (1) holds for $m = 1$ and $m = 2$ for all $\nu \geqslant 4$. To obtain adequately sharp estimates we also verify (1) directly for $m = 3$. It suffices to show that

$$\sup_{\nu \geqslant 4} \frac{\left(\frac{7}{36}\right)^\nu + \left(\frac{21}{36}\right)^\nu + \left(\frac{35}{36}\right)^\nu}{\left(\frac{1}{3}\right)^\nu + \left(\frac{2}{3}\right)^\nu + 1} < \frac{6}{7}.$$

We do this by breaking $[4, \infty)$ into a finite number of intervals $[a, b)$ on each of which we have

$$\sup_{a \leqslant \nu < b} \frac{\left(\frac{7}{36}\right)^\nu + \left(\frac{21}{36}\right)^\nu + \left(\frac{35}{36}\right)^\nu}{\left(\frac{1}{3}\right)^\nu + \left(\frac{2}{3}\right)^\nu + 1} < \frac{\left(\frac{7}{36}\right)^a + \left(\frac{21}{36}\right)^a + \left(\frac{35}{36}\right)^a}{\left(\frac{1}{3}\right)^b + \left(\frac{2}{3}\right)^b + 1} < \frac{6}{7}.$$

One set of such intervals has end points at $4, 4.3, 4.7, 5.3, 6.4, 11, \infty$.

Suppose now that (3) holds for some $m \geqslant 3$. Then by our inductive assumption

$$\sum_{j=1}^{2m+2} j^\nu < \left\{1 + \left(\frac{2m}{2m+1}\right)^{\nu+1}\right\} 2^\nu \sum_{j=1}^{m} j^\nu + (2m+1)^\nu + (2m+2)^\nu,$$

and we must show that the last expression is at most

$$\left\{1 + \left(\frac{2m+2}{2m+3}\right)^{\nu+1}\right\} 2^\nu \sum_{j=1}^{m+1} j^\nu.$$

Equivalently, we must show that

$$\begin{aligned} &\left(m + \frac{1}{2}\right)^\nu - \left(\frac{2m+2}{2m+3}\right)^{\nu+1} (m+1)^\nu \\ &\quad < \left\{\left(\frac{2m+2}{2m+3}\right)^{\nu+1} - \left(\frac{2m}{2m+1}\right)^{\nu+1}\right\} \sum_1^m j^\nu. \end{aligned} \tag{4}$$

We give an upper estimate for the left side of (4) and lower estimates for each of the factors on the right side of (4).

LEMMA 1. *Let A denote the left-hand side of* (4). *Then*

$$A < \frac{(m+1/2)^\nu}{2m+3}\left\{1 - \frac{\nu}{2m+2}\right\} \tag{5}$$

for any positive ν.

Proof. We have

$$\begin{aligned}\frac{A}{(m+1/2)^\nu} &= 1 - \left\{\frac{(2m+1)(2m+3)}{(2m+2)(2m+2)}\right\}^{-\nu}\frac{2m+2}{2m+3}\\ &= 1 - \left\{1 - \frac{1}{(2m+2)^2}\right\}^{-\nu}\frac{2m+2}{2m+3}\\ &< 1 - \left\{1 + \frac{\nu}{(2m+2)^2}\right\}\frac{2m+2}{2m+3} = \frac{1}{2m+3} - \frac{\nu}{(2m+2)(2m+3)}.\end{aligned}$$

LEMMA 2. *Suppose that $3 \leqslant \nu \leqslant 2m+1$. Then*

$$\sum_{n=1}^{m} n^\nu > \frac{(m+1/2)^{\nu+1}}{\nu+1}\left\{1 - \frac{2\nu(\nu+1)}{27(m+1)^2}\right\}. \tag{6}$$

Proof. We have by two integrations by parts

$$\begin{aligned}&\int_{n-1/2}^{n} t^\nu\,dt + \int_n^{n+1/2} t^\nu\,dt - n^\nu\\ &\quad= \frac{\nu(\nu-1)}{2}\left\{\int_{n-1/2}^{n}(t-n+1/2)^2 t^{\nu-2}dt\right.\\ &\qquad\qquad\left. + \int_n^{n+1/2}(t-n-1/2)^2 t^{\nu-2}\,dt\right\}\\ &\quad= \frac{\nu(\nu-1)}{2}\int_0^{1/2}(u-1/2)^2\left\{(n-u)^{\nu-2} + (n+u)^{\nu-2}\right\}du\\ &\quad\leqslant \frac{\nu(\nu-1)}{2}\left\{(n-1/2)^{\nu-2} + (n+1/2)^{\nu-2}\right\}\int_0^{1/2}(u-1/2)^2\,du,\end{aligned}$$

where we have used the fact that $(n-u)^{\nu-2} + (n+u)^{\nu-2}$ is an increasing function of u on $[0, 1/2]$. By addition it follows that

$$\int_{1/2}^{m+1/2} t^\nu\,dt \leqslant \sum_{n=1}^{m} n^\nu + \frac{\nu(\nu-1)}{48}\sum_{n=1}^{m}\left\{(n-1/2)^{\nu-2} + (n+1/2)^{\nu-2}\right\}.$$

By convexity

$$y^{\nu-2} \leqslant \int_{y-1/2}^{y+1/2} x^{\nu-2}\,dx \qquad (y \geqslant 1/2),$$

so

$$\frac{(m+1/2)^{\nu+1}-(1/2)^{\nu+1}}{\nu+1} \leqslant \sum_{n=1}^{m} n^{\nu} + \frac{\nu(\nu-1)}{48}\left\{\int_0^m t^{\nu-2}\,dt + \int_1^{m+1} t^{\nu-2}\,dt\right\}.$$

If we evaluate the last integrals and note that $2^{-\nu-1}/(\nu+1) - \nu/48 < 0$, we obtain

$$\frac{(m+1/2)^{\nu+1}}{\nu+1} < \sum_{n=1}^{m} n^{\nu} + \frac{\nu}{48}(m+1/2)^{\nu-1}\left\{\left(\frac{2m}{2m+1}\right)^{\nu-1} + \left(\frac{2m+2}{2m+1}\right)^{\nu-1}\right\}.$$

For fixed m, the expression in braces is an increasing function of ν. For $\nu \leqslant 2m+1$ it follows that

$$\frac{(m+1/2)^{\nu+1}}{\nu+1} < \sum_{n=1}^{m} n^{\nu} + \frac{\nu}{48}(m+1/2)^{\nu-1}\left\{\left(\frac{2m}{2m+1}\right)^{2m} + \left(\frac{2m+2}{2m+1}\right)^{2m}\right\}.$$

We complete the proof of the lemma by showing that

$$\left(\frac{2m}{2m+1}\right)^{2m} + \left(\frac{2m+2}{2m+1}\right)^{2m} < \frac{8}{9}\left(\frac{2m+1}{m+1}\right)^2, \qquad m \geqslant 3. \tag{7}$$

This relation is verified directly for $m = 3, 4, 5, 6$. For $m \geqslant 7$ we have

$$\begin{aligned}\left(\frac{2m}{2m+1}\right)^{2m} + \left(\frac{2m+2}{2m+1}\right)^{2m} &< \left(\frac{2m}{2m+1}\right)^{2m+1} + \left(\frac{2m+2}{2m+1}\right)^{2m+1}\\ &< \frac{1}{e} + e < \frac{25}{8} \leqslant \frac{8}{9}\left(\frac{2m+1}{m+1}\right)^2.\end{aligned}$$

Now (7) fails for $m = 1, 2$, but (6) holds by direct calculation for $m = 1$, $\nu = 3$ and for $m = 2$, $3 \leqslant \nu \leqslant 5$.

The latter verification can be accomplished by the aforementioned device of breaking $[3, 5]$ into a finite number of subintervals.

LEMMA 3. *Suppose that* $4 \leqslant \nu \leqslant 2m$ *and* $m \geqslant 3$. *Then*

$$\left(\frac{2m+2}{2m+3}\right)^{\nu+1} - \left(\frac{2m}{2m+1}\right)^{\nu+1} > \frac{2(\nu+1)B}{(2m+1)(2m+3)},$$

where

$$B > 1 - \frac{\nu/2}{m+1} + \frac{\nu(\nu-1)}{8(m+1)^2} - \frac{\nu^2(\nu-1)}{48(m+1)^3}.$$

Proof. Let

$$f(x) = x^{\nu+1}, \qquad \frac{2m}{2m+1} = a, \qquad \frac{2m+2}{2m+3} = b,$$

and let $c = (a+b)/2$. Noting that $f''' \geqslant 0$ on $[a, b]$, we have by Taylor's formula

$$f(b) = f(c) + (b-c)f'(c) + (b-c)^2 f''(c)/2 + (b-c)^3 f'''(\xi)/6,$$
$$f(a) = f(c) + (a-c)f'(c) + (a-c)^2 f''(c)/2 + (a-c)^3 f'''(\eta)/6.$$

We obtain, upon subtracting,

$$f(b) - f(a) > (b-a)f'(c) = \frac{2}{(2m+1)(2m+3)}(\nu+1)B,$$

where

$$B^{1/\nu} = \frac{4m^2+6m+1}{(2m+1)(2m+3)} = 1 - \frac{2m+2}{(2m+1)(2m+3)}.$$

Let

$$\varepsilon = \frac{2m+2}{(2m+1)(2m+3)}.$$

Then

$$\varepsilon = \frac{1}{2m+2} + \frac{1}{(2m+2)^3} + \frac{1}{(2m+1)(2m+2)^3(2m+3)} = \varepsilon_1 + \varepsilon_2 + \varepsilon_3,$$

say. By the binomial theorem

$$B = (1-\varepsilon)^\nu > 1 - \nu\varepsilon + \nu(\nu-1)\varepsilon^2/2 - \nu(\nu-1)(\nu-2)\varepsilon^3/6,$$

where the truncation estimate is valid since $\nu\varepsilon < 1$.

Let $S_1 = 1 - \nu\varepsilon + \nu(\nu-1)\varepsilon^2/4$. Then

$$S_1 > 1 - \nu(\varepsilon_1 + \varepsilon_2) + \nu(\nu-1)\varepsilon_1^2/4.$$

Let $S_2(t) = t^2/2 - (\nu-2)t^3/3$. Since $S_2(t)$ is increasing for $0 < t \leqslant \varepsilon$, we have $S_2(\varepsilon) > S_2(\varepsilon_1)$.

Thus

$$B > S_1 + \frac{\nu(\nu-1)}{2}S_2(\varepsilon_1)$$
$$> 1 - \frac{\nu}{2m+2} - \frac{\nu}{(2m+2)^3} + \frac{\nu(\nu-1)}{2(2m+2)^2} - \frac{\nu(\nu-1)(\nu-2)}{6(2m+2)^3}.$$

Finally, for $\nu \geqslant 4$ we have

$$\frac{-\nu - \nu(\nu-1)(\nu-2)/6}{(2m+2)^3} \geqslant \frac{-\nu^2(\nu-1)/6}{(2m+2)^3},$$

which completes the estimate of B.

We now use Lemmas 1, 2, and 3 to establish the validity of (4). By these lemmas it suffices to show that

$$\frac{(m+1/2)^{\nu}}{2m+3}\left\{1-\frac{\nu}{2m+2}\right\}$$
$$<\frac{2(\nu+1)}{(2m+3)(2m+1)}\left\{1-\frac{\nu}{2m+2}+\frac{\nu(\nu-1)}{8(m+1)^2}-\frac{\nu^2(\nu-1)}{48(m+1)^3}\right\}$$
$$\cdot\frac{(m+1/2)^{\nu+1}}{\nu+1}\left\{1-\frac{2\nu(\nu+1)}{27(m+1)^2}\right\}$$

for $4 \leqslant \nu < 2m$, $m \geqslant 3$. Equivalently, we must show that

$$1-\frac{\nu}{2m+2}<\left\{1-\frac{\nu}{2m+2}+\frac{\nu(\nu-1)}{8(m+1)^2}-\frac{\nu^2(\nu-1)}{48(m+1)^3}\right\}\left\{1-\frac{2\nu(\nu+1)}{27(m+1)^2}\right\},$$

or that

$$\frac{(11\nu-43)\nu}{216(m+1)^2}+\frac{(21\nu+75)\nu^2}{1296(m+1)^3}-\frac{\nu^2(\nu-1)(\nu+1)}{108(m+1)^4}+\frac{(\nu-1)(\nu+1)\nu^3}{648(m+1)^5}>0.$$

The first term in the last expression is positive for $\nu \geqslant 4$. The second plus the fourth terms exceed the third term, by the arithmetic-geometric mean inequality. This establishes (4) and completes proof of the induction.

An asymptotic estimate.

LEMMA 4. *If* $\nu \in \mathbb{R}$, $\nu \neq -1$, *then*

$$\sum_{j=1}^{m} j^{\nu}=\frac{m^{\nu+1}}{\nu+1}+\frac{1}{2}m^{\nu}+\frac{\nu}{12}m^{\nu-1}+\zeta(-\nu)+O(m^{\nu-3}),$$

where ζ *denotes the Riemann zeta function.*

This is a particular case of the Euler-Maclaurin sum formula and may be found in B. C. Berndt's *Ramanujan's Notebooks*, Part I, Springer-Verlag, New York, 1985, p. 150.

If $\nu > -1$, then two applications of Lemma 4 and a division yield

$$\begin{aligned}\sum_{j=1}^{2m} j^{\nu}\Big/\sum_{j=1}^{m}(2j)^{\nu}&=2-\frac{\nu+1}{2m}+\frac{(\nu+1)(\nu+2)}{8m^2}-\frac{(\nu+1)^2(\nu+6)}{48m^3}\\&\quad-\frac{(\nu+1)\zeta(-\nu)(2^{\nu+1}-1)}{2^{\nu}m^{\nu+1}}+O\left\{\frac{1}{m^{\nu+2}}+\frac{1}{m^{2\nu+2}}+\frac{1}{m^4}\right\}.\end{aligned} \tag{8}$$

For fixed ν, the binomial theorem gives

$$1+\left(\frac{2m}{2m+1}\right)^{\nu+1}=1+\left(1+\frac{1}{2m}\right)^{-\nu-1}$$
$$=2-\frac{\nu+1}{2m}+\frac{(\nu+1)(\nu+2)}{8m^2}-\frac{(\nu+1)(\nu+2)(\nu+3)}{48m^3}$$
$$+O\left(\frac{1}{m^4}\right). \tag{9}$$

If we subtract (8) from (9) we obtain, for $\nu > -1$,

$$\Delta_\nu(m)=1+\left(\frac{2m}{2m+1}\right)^{\nu+1}-\sum_{j=1}^{2m} j^\nu \Big/ \sum_{j=1}^{m}(2j)^\nu$$
$$=\frac{\nu(\nu+1)}{24m^3}+\frac{(\nu+1)\zeta(-\nu)(2^{\nu+1}-1)}{2^\nu m^{\nu+1}}+O\left(\frac{1}{m^{\nu+2}}+\frac{1}{m^{2\nu+2}}+\frac{1}{m^4}\right).$$

For $\nu \leqslant -1$ we estimate $\Delta_\nu(m)$ by approximating $\sum_{j=1}^{2m} j^\nu$ and $\sum_{j=1}^{m}(2j)^\nu$ directly. Together these estimates give

THEOREM 2. *As* $m \to \infty$

$$\Delta_\nu(m) \sim \nu(\nu+1)/\{24m^3\}, \qquad \text{if} \quad \nu \geqslant 2,$$
$$\Delta_\nu(m) \sim \frac{(\nu+1)(2^{\nu+1}-1)}{2^\nu m^{\nu+1}}\zeta(-\nu), \quad \text{if} \quad -1<\nu<2,$$
$$\Delta_{-1}(m) \sim \frac{-2\log 2}{\log m},$$
$$\Delta_\nu(m) \sim 2-2^{-\nu}, \qquad \text{if} \quad \nu<-1.$$

In particular, $\Delta_\nu(m)$ *is asymptotically positive if and only if* $\nu \geqslant 2$.

A Product Inequality

E 3132 [1986, 132]. *Proposed by Robert E. Shafer, Berkeley, CA.*

For integers $n \geqslant 2$ and real $s > 0$, show that

$$\left(\prod_{i=0}^{n}(s+i)\right)\left(\sum_{j=0}^{n}\frac{1}{s+j}\right)<(n+1)\prod_{k=1}^{n}\left(s+k-\frac{1}{2}\right).$$

Solution I by San Bernardino Problem Solving Group, San Bernardino, CA. By defining $f(s)=\prod_{k=0}^{n}(s+k)$, the desired inequality becomes $f'(s)<f(s+\frac{1}{2})-f(s-\frac{1}{2})$. Writing f as a Taylor series about s yields

$$f\left(s+\frac{1}{2}\right)-f\left(s-\frac{1}{2}\right)=f'(s)+\sum f^{(j)}(s)\left(\frac{1}{2}\right)^{j-1}\Big/ j!,$$

where the summation is over odd values of j from 3 to $n+1$. For $n \geqslant 2$ and $s > 0$, the higher order derivatives are positive (f is a polynomial with positive coefficients), which establishes the inequality.

Solution II by Michael Vowe, Therwil, Switzerland. First we prove by induction the inequality

$$\prod_{i=0}^{n}(s+i) < \left(s+\frac{n}{2}\right)\prod_{k=1}^{n}\left(s+k-\frac{1}{2}\right). \tag{1}$$

For $n=1$, this is $s(s+1) < s^2+s+\frac{1}{4} = (s+\frac{1}{2})^2$. For the induction step, it suffices to show

$$(s+n/2)(s+n+1) < \left(s+\frac{n+1}{2}\right)\left(s+n+\frac{1}{2}\right),$$

which reduces to $n+(1/2) > n$.

Now we prove the original inequality by induction. For $n=2$, the inequality becomes

$$s(s+1)(s+2)\left(\frac{1}{s}+\frac{1}{s+1}+\frac{1}{s+2}\right) = 3s^2+6s+2 < 3s^2+6s+\frac{9}{4}$$
$$= 3\left(s+\frac{1}{2}\right)\left(s+\frac{3}{2}\right).$$

Now assume that $n \geqslant 2$ and that the original inequality holds for n. By the induction hypothesis and (1) we have

$$\prod_{i=0}^{n+1}(s+i)\sum_{j=0}^{n+1}\frac{1}{s+j} = \prod_{i=0}^{n+1}(s+i)\sum_{j=0}^{n}\frac{1}{s+j} + \prod_{i=0}^{n}(s+i)$$
$$< (s+n+1)(n+1)\prod_{k=1}^{n}\left(s+k-\frac{1}{2}\right)$$
$$+\left(s+\frac{n}{2}\right)\prod_{k=1}^{n}\left(s+k-\frac{1}{2}\right)$$
$$= (n+2)\prod_{k=1}^{n+1}\left(s+k-\frac{1}{2}\right).$$

Solution III and generalization by Behdžet A. Mesihović, University Džemal Bijedić, Mostar, Yugoslavia. For a sequence $\langle a_i\rangle$ of positive numbers, the Arithmetic-Harmonic Mean Inequality is

$$A_n(a) = \frac{1}{n}\sum_{i=1}^{n}a_i \geqslant \frac{n}{\sum_{i=1}^{n}a_i^{-1}} = H_n(a),$$

with equality if and only if $a_1 = \cdots = a_n$. With this, we can generalize the original

claim to

$$(n+1)\prod_{k=1}^{n}\frac{a_k+a_{k+1}}{2} \geqslant \left(\sum_{i=1}^{n+1} a_i^{-1}\right)\left(\prod_{j=1}^{n+1} a_j\right) \tag{2}$$

for a sequence of positive numbers satisfying

$$(k+1)a_k-(k-1)a_{k+1} \geqslant 2A_k(a), \tag{3}$$

with equality in (2) if and only if $a_1 = \cdots = a_{n+1}$ or $n = 1$. Note that the condition (3) holds for any concave sequence, and that if $a_k = s + k - 1$, then it holds with equality.

If $n = 1$, then (2) holds with equality. For $n > 1$, we apply induction. Letting L_n, R_n denote the left side and right side of (2), we rewrite $2n(L_n - R_n)$ and then use the induction hypothesis to obtain

$$\begin{aligned}2n(L_n-R_n) &= (n+1)(a_n+a_{n+1})L_{n-1} - 2n\left(a_{n+1}R_{n-1} + \prod_{k=1}^{n} a_k\right)\\ &\geqslant \left((n+1)a_n-(n-1)a_{n+1}\right)R_{n-1} - 2n\prod_{k=1}^{n} a_k\\ &\geqslant 2A_n(a)R_{n-1} - 2H_n(a)R_{n-1} \geqslant 0.\end{aligned}$$

The final inequality is strict unless $a_1 = \cdots = a_n$, which completes the proof of (2).

Also solved by 25 other readers and the proposer.

ADVANCED PROBLEMS

For instructions about submitting solutions of these Advanced Problems, which should be mailed before June 30, 1988, see the inside front cover. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label is an acknowledgement is desired.

6566. *Proposed by Jonathan E. Spingarn, Georgia Institute of Technology.*

Let r: $R^2 \to R^2$ be a rotation about the origin by an irrational multiple of π. Let f: $R^2 \to R^2$ be defined by

$$f(x,y) = (x,y) \quad \text{if} \quad x \le 1, \qquad f(x,y) = (2-x,y) \quad \text{if} \quad x > 1.$$

Does there exist $(x_0, y_0) \in R^2$ such that

$$\left\|(r\circ f)^k(x_0,y_0)\right\| > 1$$

for all positive integers k? (Here $(r \circ f)^k$ denotes the kth iterate of $r \circ f$ and $\|\ \|$ denotes the ordinary Euclidean norm.)

6567. *Proposed by Gerald A. Edgar, Ohio State University, and Lee A. Rubel, University of Illinois at Urbana-Champaign.*

Let us say that a function $u(x, y)$ is algebraic if there exists a nontrivial polynomial P in three variables such that $P(x, y, u(x, y)) = 0$. Suppose $u(x, y)$ is

a harmonic and algebraic function on a region G in $\mathbb{R}^2$. Must its harmonic conjugate v also be algebraic?

SOLUTIONS OF ADVANCED PROBLEMS

Monotonic Norms

6511 [1986, 218]. *Proposed by M. Luisa N. McAllister, Moravian College, Bethlehem, PA.*

Let $R^{m\times n}$ be the vector space of all $m \times n$ matrices with real entries. Suppose that $A = (a_{ij})$, $B = (b_{ij})$ are in $R^{m\times n}$ and satisfy $0 \leqslant a_{ij} \leqslant b_{ij}$ for all i, j. For which norms ν on $R^{m\times n}$ does the inequality

$$0 \leqslant \nu(A) \leqslant \nu(B)$$

always hold?

Editorial Comment. No solutions were received. The problem essentially asks which norms $F(x_{11}, x_{12}, \ldots, x_{mn})$ in mn-dimensional Euclidean space are non-decreasing functions of each coordinate separately for nonnegative values of $x_{11}, x_{12}, \ldots, x_{mn}$.

The max norm, the max row-sum norm, the max column-sum norm, and all p-norms with $p > 0$ certainly have this monotonicity property. The square root of a positive definite quadratic form in $x_{11}, x_{12}, \ldots, x_{mn}$ has this monotonicity property if and only if all the coefficients of the form are nonnegative. For example

$$\left\{x_{11}^2 + x_{12}^2 + \cdots + x_{mn}^2 + \varepsilon(x_{11} - x_{12})^2\right\}^{1/2}$$

does not have this monotonicity property for any positive ε.

6512 [1986, 218]. *Proposed by Alfonso Villani, Universita di Catania, Italy.*

A Measured Response

Let (Ω, S, μ) be a positive measure space. Let A be a Borel subset of $\mathbb{C}$ (the complexes) and let $M(A)$ denote the subset of $L^1(\mu)$ defined by

$$M(A) = \left\{f \in L^1(\mu) : \mu\left(f^{-1}(A)\right) > 0\right\}.$$

(a) Assume that A is a set of first category not containing zero. Prove that $M(A)$ is of first category.

(b) Assume that A is of second category. Prove that $M(A)$ is of second category.

(c) For which measures μ is the assumption "$0 \notin A$" in (a) superfluous?

Solution by Kenneth Schilling, University of Michigan, Flint. We first establish three lemmas.

LEMMA 1. *Suppose $A \subseteq \mathbb{C}$ is closed, $0 \notin A$, and $\varepsilon > 0$; then the set*

$$\left\{f \in L^1(\mu) : \mu\left(f^{-1}(A)\right) < \varepsilon\right\}$$

is open.

Proof. Fix $f \in L^1(\mu)$ such that $\mu(f^{-1}(A)) < \varepsilon$. We will find $\delta > 0$ such that

$$N_\delta(f) = \left\{g \in L^1(\mu):\ \int_\Omega |f - g|\,d\mu < \delta\right\} \subseteq \{f : \mu(f^{-1}(A)) < \varepsilon\}.$$

Let

$$\Omega_k = \{\omega \in \Omega:\ d(f(\omega), A) \leqslant 1/k\}$$

where $d(x, A)$ denotes the Euclidean distance from the point x to the set A. Since A is closed,

$$f^{-1}(A) = \bigcap_{k=1}^{\infty} \Omega_k.$$

Since A is bounded away from 0, $\mu(\Omega_k) < \infty$ for all sufficiently large k. Therefore,

$$\mu(\Omega_k) = \varepsilon' < \varepsilon$$

for some k. Choose $\delta = (\varepsilon - \varepsilon')/k$. Then for all $g \in N_\delta(f)$ we have

$$\begin{aligned}(\varepsilon - \varepsilon')/k > \int |f - g|\,d\mu &\geqslant \int_{g^{-1}(A)\setminus\Omega_k} |f - g|\,d\mu \\ &\geqslant [\mu(g^{-1}(A)) - \varepsilon']/k,\end{aligned}$$

so $\mu(g^{-1}(A)) < \varepsilon$ as desired.

LEMMA 2. *If $A \subset \mathbf{C}$ is open and $\varepsilon > 0$, then*

$$\{f \in L^1(\mu):\ (f^{-1}(A)) > \varepsilon\}$$

is open.

The proof of Lemma 2 (omitted) is similar to that of Lemma 1.

LEMMA 3. *If $U \subset \mathbf{C}$ is open, $0 \notin U$, and $A \subseteq U$ is dense in U, then the set*

$$\{f \in L^1(\mu) : f(\Omega) \cap U \subseteq A\}$$

is dense in $L^1(\mu)$.

Proof. Fix $g \in L^1(\mu)$, and $\varepsilon > 0$. We will construct an $f \in L^1(\mu)$ such that

$$f(\Omega) \cap U \subseteq A \quad \text{and} \quad \int |f - g|\,d\mu < \varepsilon.$$

First, let a_i, $i = 1, 2, 3, \ldots$, be a sequence of elements of A that is dense in U. Next, define a function

$$h : \mathbf{C} \to \mathbf{C}$$

as follows. If $z \notin U$, set $h(z) = z$. If $z \in U$, let $h(z) = a_i$ where i is the least integer such that

$$|a_i - z| \leqslant \varepsilon |z| \Big/ \int |g|\,d\mu.$$

(It is here that we use the fact that $0 \notin U$.) Then h is Borel-measurable, $h(U) \subseteq A$, and

$$|z - h(z)| \leqslant \varepsilon|z| \Big/ \int |g|\,d\mu$$

for all $z \in \mathbb{C}$. Finally, let $f = h \circ g$ be the composition of h with g. By the above, f is μ-measurable,

$$f(\Omega) \cap U \subseteq A,$$

and

$$\int |f - g|\,d\mu = \int |h \circ g - g|\,d\mu$$
$$\leqslant \int \left(\varepsilon|g| \Big/ \int |g|\,d\mu \right) d\mu = \varepsilon.$$

This proves Lemma 3, and now we can solve the problem.

Proof of (a). If A is of the first category, then

$$A \subseteq \bigcup_n C_n,$$

where each C_n is closed and nowhere dense, and $0 \notin C_n$. Since

$$M(A) \subseteq \bigcup_n M(C_n),$$

it suffices to show that $M(C_n)$ is of the first category, for each n. By Lemma 1,

$$L^1(\mu)/M(C_n) = \left\{ f \in L^1(\mu) : \mu\left(f^{-1}(C_n)\right) = 0 \right\}$$
$$= \bigcap_{m=1}^{\infty} \left\{ f : \mu\left(\cdot f^{-1}(C_n)\right) < 1/m \right\} \qquad (*)$$

is a G_δ set. Also, by Lemma 3 with

$$U = \mathbb{C} \setminus \{0\} \quad \text{and} \quad A = \mathbb{C} \setminus C_n$$

(this is dense in $\mathbb{C}$), we see that

$$\left\{ f \in L^1(\mu) : f(\Omega) \setminus \{0\} \subseteq \mathbb{C} \setminus C_n \right\}$$
$$= \left\{ f : f(\Omega) \subseteq \mathbb{C} \setminus C_n \right\}$$

(note that $0 \in \mathbb{C} \setminus C_n$) is dense. Since it is also contained in

$$L^1(\mu) \setminus M(C_n)$$

we conclude that $M(C_n)$ is of the first category.

Assertion (b) is incorrect if Ω has no subset of finite, positive measure. For if we let $A = \mathbb{C} \setminus \{0\}$, then A is of the second category, but $M(A) = \varnothing$.

However, (b) is true provided there exists a measurable set $X \subseteq \Omega$ such that $0 < \mu(X) < \infty$. To prove this, suppose that A is not of the first category. Then there exists an open set $U \subseteq \mathbb{C}$ such that 0 is not in the closure of U, and $U \setminus A$ is of the first category. Therefore, there exist open sets $B_1 \supseteq B_2 \supseteq B_3 \supseteq \cdots$ such that $U \supseteq B_1$, each B_n is dense in U, and

$$\bigcap_{n=1}^{\infty} B_n \subseteq A.$$

Choose ε such that $0 < \varepsilon < \mu(X)$. By Lemma 2, the set

$$F_n := \{f : \mu(f^{-1}(B_n)) > \varepsilon\}$$

is open for each n. Since the assertions

$$\varepsilon < \mu(f^{-1}(U)) \quad \text{and} \quad f(\Omega) \cap U \subseteq B_n$$

together imply that $\mu(f^{-1}(B_n)) > \varepsilon$, it follows by Lemma 3 that F_n is dense in the nonempty open set

$$Q := \{f : \mu(f^{-1}(U)) > \varepsilon\}.$$

(To prove that Q is nonempty, choose $x_0 \in U$ and let $f(\omega) = x_0$ for $\omega \in X$, and 0 otherwise. Then $f \in Q$.) Hence, $\bigcap_n F_n$ is the complement of a set that is of the first category in Q. Hence, it is not of the first category itself. Finally, since each $B_n \subseteq U$ is by construction bounded away from 0, for all $f \in L^1(\mu)$ we have

$$\mu(f^{-1}(B_n)) < \infty.$$

Therefore,

$$\bigcap_n F_n \subseteq \left\{f : \mu\left(f^{-1}\left(\bigcap_n B_n\right)\right) \geqslant \varepsilon\right\} \subseteq M\left(\bigcap_n B_n\right) \subseteq M(A),$$

so $M(A)$ is not of the first category.

Solution to (c). The condition $0 \notin A$ in (a) is superfluous if and only if the measure μ is σ-finite.

Suppose that μ is not σ-finite. From the well-known fact that, for all $f \in L^1(\mu)$, the set

$$\Omega \setminus f^{-1}(\{0\})$$

is of σ-finite μ-measure, it follows that $M(\{0\}) = L^1(\mu)$. Of course, $\{0\}$ is of the first category and $L^1(\mu)$ is not. On the other hand, suppose that μ is a σ-finite measure. In order to eliminate the hypothesis "$0 \notin A$" from (a), we shall (1) reformulate Lemma 1 so as to avoid using this hypothesis at line (*) of the proof of (a), and (2) show how to eliminate the condition "$0 \notin U$" from Lemma 3. In 1 and 2 below, let $\Omega = \cup_n \Omega_n$, where the sets Ω_n are disjoint and of finite μ-measure.

1. Lemma 1 as stated is false if the hypothesis "$0 \notin A$" is omitted, but a proof similar to the one given proves that if $A \subseteq \mathbb{C}$ is closed, $\varepsilon > 0$, and $\Omega' \subseteq \Omega$ is a set of finite μ-measure, then the set

$$\{ f \in L^1(\mu) : \mu(\Omega' \cap f^{-1}(A)) < \varepsilon \}$$

is open. Now in the proof of (a), replace the lines immediately following (*) with

$$\text{“}L^1(\mu) \setminus M(C_n) = \{ f \in L^1(\mu) : \mu(f^{-1}(C_n)) = 0 \}$$
$$= \bigcap_{k=1}^{\infty} \bigcap_{m=1}^{\infty} \{ f : \mu(\Omega_k \cap f^{-1}(C_n)) < 1/m \} \text{ is a } G_\delta \text{ set.”}$$

2. To eliminate the condition "$0 \notin U$" from Lemma 3, replace the single function $h(z)$ in the proof of Lemma 3 by a sequence of functions.

For $n = 1, 2, 3, \ldots$ define

$$h_n : \mathbb{C} \to \mathbb{C}$$

as follows. If $z \notin U$, set $h_n(z) = z$. If $z \in U$, let $h_n(z) = a_i$ where i is the least integer such that

$$|a_i - z| < \varepsilon \cdot 2^{-n}/\mu(\Omega_n).$$

Finally, define $f : \Omega \to \mathbb{C}$ by $f(\omega) = h_n(g(\omega))$, where $\omega \in \Omega_n$. Thus,

$$\int_{\Omega_n} |f - g| \, d\mu = \int_{\Omega_n} |h_n \circ g - g| \, d\mu$$
$$< \int_{\Omega_n} \varepsilon 2^{-n}/\mu(\Omega_n) \, d\mu = \varepsilon 2^{-n}.$$

Summation over n yields the conclusion of Lemma 3, as desired.

Also solved by Victor Pambuccian (Romania) and the proposer.

Choice on Squares

6518 [1986, 403]. *Proposed by Vladimir N. Akis, California State University at Los Angeles.*

Let S be a closed square of area 4. Denote by $\mathscr{C}$ the collection of all squares of area 1 contained in S whose centers are points of S, and whose sides are parallel to the sides of S.

Let $f : \mathscr{C} \to S$ be a choice function, where $f(A) \in A$ for each $A \in \mathscr{C}$. Suppose that f is continuous with respect to the Hausdorff metric on $\mathscr{C}$. Show that $f(A)$ is the center of S for some $A \in \mathscr{C}$.

Solution by David C. Kay, University of North Carolina, Asheville. Let c denote the center of S. Suppose $f(A) \neq c$ for all $A \in \mathscr{C}$. Then no sequence $\{A_i\}$ of sets in

$\mathscr{C}$ can satisfy

$$\lim f(A_i) = c.$$

Were this to happen, some subsequence $\{B_j\}$ of $\{A_i\}$ would converge in the Hausdorff metric to a square $B \in \mathscr{C}$, and then by continuity of f, we would have

$$f(B) = f(\lim B_j) = \lim f(B_j) = c,$$

a contradiction. Hence there exists a disc D of positive radius, centered at c, such that $f(A) \notin D$, all $A \in \mathscr{C}$. But this allows us to construct a retraction g of D to its boundary defined as follows: For each $x \in D$ consider the line segment joining x and $f(A_x)$, where A_x is the unique square in $\mathscr{C}$ centered at x. This line segment meets the circular boundary of D at a unique point $g(x)$. Clearly $g(x)$ is continuous. But this contradicts the No Retraction Theorem.

All other solutions were rather similar to the above, and relied on either the Brouwer fixed point theorem or some closely related result. The solutions of John A. Frohliger and John Henry Steelman both extended the result to n dimensions. In addition, Frohliger remarks that if the square S is replaced by the closed disc of radius 2 and $\mathscr{C}$ by the internal closed discs of radius r (where $r < 1$ is fixed), then the range of f even contains a centered closed disc of radius $2 - 2r$.

Also solved by G. D. Chakerian, William Geller, Victor Pambuccian (Romania), and the proposer.

A Nonhypergeometric Sum

6519 [1986, 403]. *Proposed by Ira Gessel, Brandeis University.*

Let

$$F(a, b, m, n) = \sum_{j=0}^{m} \binom{a+m+n-2j}{n-j} \binom{a+n}{j} \binom{b+m}{m-j},$$

where m and n are nonnegative integers. Show that $F(a, b, m, n) = F(b, a, n, m)$.

Solution I by Jay Hook, Champaign, Illinois. Let X and Y be two disjoint sets with

$$|X| = a + n, \quad \text{and} \quad |Y| = b + m.$$

Consider the class of all partitions of each of X and Y into three disjoint subsets, say,

$$X = X_1 \cup X_2 \cup X_3 \quad \text{and} \quad Y = Y_1 \cup Y_2 \cup Y_3,$$

such that

(i) $$|Y_1| = |X_1| + b$$

and

(ii) $$|X_2| = |Y_2| + a.$$

We refer to this class as the set of bipartitions of X and Y with parameters

(a, b, m, n). All such partitions may be constructed by the following four-step procedure:

(1) Choose a nonnegative integer j to be the cardinality of X_1. By (i) we have $j \leqslant m$.

(2) Choose X_1, a subset of X. This can be done in

$$\binom{a+n}{j}$$

different ways.

(3) Choose Y_1, a subset of Y, so that (i) is satisfied. This can be done in

$$\binom{b+m}{j+b} = \binom{b+m}{m-j}$$

different ways.

(4) Observe that the sets X_2, X_3, Y_2, Y_3 are uniquely determined by specifying the set

$$S = X_3 \cup Y_2 \subseteq (X \cup Y) \setminus (X_1 \cup Y_1).$$

Since

$$\begin{aligned} |S| &= |X| - |X_1| - |X_2| + |Y_2| \\ &= (a+n) - j + (|Y_2| - |X_2|), \end{aligned}$$

condition (ii) is equivalent to $|S| = n - j$. Hence the number of different ways of choosing an S that satisfies (ii) is

$$\binom{a+m+n-2j}{n-j}.$$

This procedure shows that the cardinality of the class of partitions satisfying (i) and (ii) is $F(a, b, m, n)$. Now clearly there is a one-to-one correspondence between the above class and the class of bipartitions of Y and X with parameters (b, a, n, m). This proves the result.

Solution II by C. L. Mallows, AT&T Bell Laboratories, Murray Hill, New Jersey. Expand the first binomial coefficient by Vandermonde's theorem to obtain

$$\binom{a+m+n-2j}{n-j} = \sum_k \binom{a+n-j}{n-j-k}\binom{m-j}{k}$$

and then rearrange the implicit factorials to obtain

$$F(a, b, m, n) = \sum_{j,k} \binom{n-k}{j}\binom{b+m-k}{m-j-k}\binom{b+m}{k}\binom{a+n}{n-k}.$$

Now (by Vandermonde again) summation over j shows the above is

$$\sum_k \binom{b+m+n-2k}{m-k}\binom{b+m}{k}\binom{a+n}{n-k} = F(b, a, n, m).$$

Proofs somewhat similar to Mallows' were also provided by Jay Hook and the proposer. According to the proposer, "unlike most binomial coefficient identities, this one is not equivalent to a well-known identity for hypergeometric series."

Reviews

Edited By Joseph Konhauser
Department of Mathematics, Macalester College, St. Paul, MN 55105

Discrete Mathematics. By John A. Dossey, Albert D. Otto, Lawrence E. Spence, and Charles Vanden Eynden. Scott, Foresman and Company, Glenview, IL 1987. 482 pp.

Richard D. Porter
Department of Mathematics, Northeastern University, Boston, MA 02115

Computers became a part of everyday life sometime between 1972, when I saw an early version of the remarkable computer film "The Hypercube" by Thomas Banchoff and Charles Strauss [**1**], and last Saturday morning when Jim Henson's Muppet Babies demonstrated the wonders of computer graphics to those who watch Saturday morning cartoons. In the Muppet show a six-year-old cartoon character, pounding away on his PC, confides to the audience that the graphics is all mathematical and very precise. At the same time, a second character runs around the screen randomly placing splotches of color on the graphics. The actions of the second character, while keeping the children's attention, suggested that precision in mathematics implies a lack of imagination.

A consequence of the pervasive use of the computer from science to word processing to Max Headroom and the Saturday morning cartoons is that students are taking courses in discrete mathematics. What is to be gained from a course in discrete mathematics? In a typical course at the freshman-sophomore level, students are looking for the background material needed for courses that follow in mathematics, computer science, engineering, etc. At the junior-senior level, students should be examining a mathematical structure in depth (group theory, for instance) and applying the results to a variety of problems. The goal of the second course is to expose students to the nature and power of mathematics.

The book under review is a text for the first of the two courses. The list of topics for such a course suggested by the MAA panel on Discrete Mathematics [**3**] includes Sets, the nature of proofs, Formal Logic, Functions and Relations, Recursion, Combinatorics, Graphs, Trees and Order, and Algebraic Structures.

The first criterion of a text for this course is that the material moves beyond definitions. Students need time to work with ideas, and see applications. A course at the freshman-sophomore level should achieve as many of the goals of the second course as possible. In the words of the MAA panel [**3**] "Adequate time should be allowed for the students to DO a lot on their own: they should be solving problems, writing proofs, constructing truth tables, manipulating symbols in Boolean algebra, deciding when, if and how to use induction, proofs by contradiction, etc." Judging from the books on my shelf, a text that covers topics in the right depth won't have space for all of the topics on the panel's list.

The book by Dossey, Otto, Spence, and Eynden passes the first test by choosing not to cover algebraic structures. The first chapter is an engaging introduction to combinatorial problems through examples of matching problems (Is it possible for an airline to meet requests from pilots for specific flights?) and Knapsack Problems (Which experiments should be packed into a satellite of given capacity in order to provide the greatest potential return to science?). The approach is one of exploring the problems. Various strategies are considered along with their complexity. The book moves on to consider a variety of problems related to graphs and develops algorithms to solve them. There are algorithms for shortest paths, searching, sorting, optimal matching, and network flows. The chapter on counting includes an algorithm for enumerating permutations. Finite State Machines, Logic and Proof Techniques are covered, but only lightly.

While the authors' choices of which topics to cover and emphasize are not right for everybody, the text succeeds in presenting mathematics in an inviting way and in greater depth than a text whose goal is to cover all the topics listed by the MAA panel. As noted by Lynn A. Steen [**2**] "The challenge for mathematicians everywhere...is to stimulate in all students the intellectual acumen that inheres in mathematical habits of thought." Discrete mathematics especially at the freshman-sophomore level, where there is a wide sampling of students, offers an excellent vehicle for meeting this challenge to the benefit of all. Hopefully, students will discover that imagination along with precision is essential in mathematics.

REFERENCES

1. "The Hypercube: Projections and Slicing" by Thomas F. Banchoff and Charles M. Strauss, 1978, International Film Bureau, Chicago, Illinois.
2. Lynn A. Steen "Power to the People," *Focus* (January-February) 1987.
3. Report: Committee on Discrete Mathematics in the First Two Years, MAA, 1986.

Constructive Analysis. (Grundlehren der mathematischen Wissenschaften; 279), by Errett Bishop and Douglas Bridges. Springer-Verlag, Berlin, Heidelberg, New York, Tokyo, 1985.

RAY MINES
Department of Mathematics, New Mexico State University, Las Cruces, NM 88003

Constructive Analysis is a substantial reworking of Errett Bishop's *Foundations of Constructive Analysis*, which greatly influenced the modern resurgence of constructive mathematics. Indeed, most of the 93 references in the new edition owe their existence to it. The effect of the book has been felt by logicians and computer scientists as well as mathematicians. Logicians have been attempting to obtain a formal system that captures Bishop's constructive framework and several computer scientists have attempted to computerize portions of the book, while others are trying to see how much mathematics they can develop under more restrictive

constructive conditions than Bishop's. Mathematicians have been proving theorems using Bishop's constructive methods. (See Stolzenberg, *Bulletin of the American Mathematical Society*, 76(1970) pp. 301–323, for a good discussion of the original.)

Bishop's book had an advantage over earlier expositions of constructive mathematics in that it was interesting and easy to read. This feature has been preserved in the new edition. Mathematicians will have no trouble understanding the material. Anyone not aware of the constructivist philosophy behind the book might wonder why the authors do some things. For example the constructive version of trichotomy is the following. If a, b and c are real numbers with $a < b$, then either $a < c$ or $c < b$.

The differences between classical mathematics and constructive mathematics stem from the different interpretations of the phrase 'there exists.' Classically, the phrase 'there exists' has come to mean 'there cannot fail to exist.' Thus most proofs of existence begin with 'assume not' and continue to arrive at a contradiction, even when it is easier to show directly that the object exists. Proofs by contradiction have become so common that they seem to be the proof of choice. Take a text off your bookshelf and open it at random. Look at the proof you find. More than likely it is a proof by contradiction. In fact, it is apt to be of the following form. First the author says assume not, then he gives a direct proof, and finally he derives a contradiction to the fact that he has a proof of what he assumed to be false.

Constructively, the phrase 'there exists' means there is an algorithm that constructs the desired object in a finite number of steps. (Here the word algorithm is undefined. It is not meant in the sense of recursive function theory where algorithm has a definite meaning. In fact several of the algorithms that are used in recursive function theory are not algorithms in our sense.) This difference in interpretation of 'there exists' is carried over to disjunction. Thus, classically to show (A or B) it is shown that $\neg A$ and $\neg B$ leads to a contradiction. Constructively one must prove either A or B.

One can best appreciate this difference by considering a proof of the Bolzano-Weierstrass Theorem. This theorem states that an infinite set S contained in the interval $[0,1]$ has a limit point. The key to the classical proof of this theorem which uses interval halving is the observation that one of the sets $S \cap [0,1/2]$ or $S \cap [1/2,1]$ must be infinite. The reason being that they cannot both be finite. Constructively this is not helpful. Knowing that they cannot both be finite does not tell us where to look for an approximation of the limit point. (Maybe this is why some students have a hard time understanding this proof: They want to know which half.) The logical principle involved here is the law of the excluded middle (for any proposition p either p or $\neg p$). In constructive mathematics the law of the excluded middle is accepted in finite situations but not in infinite ones. Thus if we know that n is an integer and n is either 0 or 1, then to show that n is 1 it is enough to show that $n \neq 0$. On the other hand if $\{a_n\}$ is a sequence of integers that are either 0 or 1 then to show that there exists an n with $a_n = 1$ we cannot argue that it is impossible for all a_n to be 0. We must instead name an integer n and show that $a_n = 1$.

In order to clarify this point let's consider the following experiment. Imagine a computer programmed to take as input an integer n and in a finite amount of time

to compute either 0 or 1 as output. Suppose further that it does it in such a way that once there is an n with output 1, then the output is 1 for all $m \geqslant n$. Thus as we feed in the sequence 1, 2,... we get a sequence $a_1 \leqslant a_2 \leqslant \cdots$ as output. This sequence has the property that given an integer n we can determine the value of a_n but we will not know the values of all the a_n's unless we find an n with $a_n = 1$, or prove that $a_n = 0$ for all n.

Suppose that you were to watch the output and you notice that $a_n = 0$ for all $n \leqslant 10^6$. You might wonder if there is an n with $a_n = 1$. As a mathematician you could examine a copy of the program and attempt to prove or disprove that $a_n = 0$ for all n. As a classical mathematician you would more than likely begin by looking for an indirect proof. Suppose you were able to give such an indirect proof to the statement that there is an n with $a_n = 1$, then your proof would have shown that it is impossible for all of the values to be 0. As a classical mathematician you would be happy as you know that there must be an n with $a_n = 1$. Now suppose that you were told that the first person to predict the value of n with $a_n = 1$ would win an all expense trip for two to Hawaii. What would you do? Your proof has not produced a method for finding such an n. If you wanted to go to Hawaii you would have to go back and attempt to find a constructive proof. This situation was nicely summarized by a colloquium speaker when he said 'This is only an existence theorem, it does not say how to get it.'

This is the situation in which the constructivist finds himself when presented with a nonconstructive proof of the existence of some mathematical object: It is nice to know it must be there, but how do we find it? Knowing that there cannot fail to exist an n with $a_n = 1$ does not allow one to do further computations. In the above situation the constructivist would try to compute a value of n and then prove that $a_n = 1$. He might not be successful because it might be a much harder task than to show that such an n cannot fail to exist.

To relate this discussion to the proof of the Bolzano-Weierstrass Theorem, let

$$S = \{b_n : b_n = 1/2 - 1/n \text{ unless } a_n = 1 \text{ in which case } b_n = 1 - 1/n\}.$$

It is clear that S is an infinite subset of the interval $[0,1]$. Which of the two sets $S \cap [0,1/2]$ and $S \cap [1/2,1]$ is infinite? To answer this question we must know if there is an n with $a_n = 1$. Notice that the sequence $\{b_n\}$ gives us an example of an increasing sequence bounded above by 1 for which we are unable to compute a limit.

You might think that the above is trivial. After all given such a program we must surely be able to answer the question on the existence of an integer n with $a_n = 1$, as a program is a finite list of instructions. The following is a simple algorithm that will produce a sequence for which we do not know if such an n exists or not. Let $a_n = 1$ if there exists an odd perfect number less than or equal to n and 0 otherwise.

An example of the problems we face when trying to prove a disjunction is given by the following classical proof of the fact that there exists two irrational numbers α and β with α^β rational. Let $\gamma = \sqrt{2}^{\sqrt{2}}$. Either γ is rational and we take $\alpha = \sqrt{2} = \beta$, or γ is irrational and we take $\alpha = \gamma$ and $\beta = \sqrt{2}$ to get $\alpha^\beta = 2$. Most mathematicians, while admiring this slick proof, would like to know which of the two pairs (γ,

$\sqrt{2}$) and ($\sqrt{2}$, $\sqrt{2}$), gives the example. A constructivist carries this wondering to such an extent that he would not consider this a proof of the stated theorem. He would need to have a proof of the fact that γ is irrational. To a constructivist the above is a proof of the following statement.

It is impossible that for each pair (α, β) *of irrational numbers the number* α^β *is irrational.*

In the Prolog on page 3 of the book under review we find the following. 'Every theorem proved with idealistic methods presents a challenge: to find a constructive version, and to give it a constructive proof.' This is what most constructive mathematicians have been doing. On the surface it looks like an easy task. Just take some well-known theorem and reprove it in a constructive way. Nevertheless, the task of finding a constructive version can be hard. Usually there are several constructive versions. This is the case for the Big Picard Theorem on essential singularities. On pages 179 and 180 we find two versions.

(6.21) THEOREM. *If f is a differentiable function on the open annulus*

$$A = \{z \in \mathbb{C} : 0 < |z - z_0| < r\}$$

with 0 and 1 not in the range, then f has a pole of determinate order at z_0.

(6.22) THEOREM. *Let f be differentiable on the open annulus*

$$A = \{z \in \mathbb{C} : 0 < |z - z_0| < r\},$$

with an essential singularity at z_0, and let ζ and ζ' be distinct complex numbers. Then in any neighborhood of z_0, f takes at least one of the values ζ and ζ'.

Classically, these two statements are equivalent. Constructively, they are quite different. Version (6.21) says that we can compute an integer n so that the coefficient of $(z - z_0)^m$ in the Laurent expansion of f is zero for all $m < n$. While in (6.22), given a neighborhood of z_0, we are able to compute a complex number z in this neighborhood for which the value of $f(z)$ is either ζ or ζ'.

At first blush the reader might think that by considering $(1/n)$-neighborhoods of z_0 we could conclude that one of the values ζ or ζ' is taken on infinitely often. All that we can prove is that it is impossible that each is taken on only a finite number of times.

Many people have expressed the view that there should be close ties between the theory of algorithms and constructive mathematics. Perhaps because algorithms are innately constructive. One concern of people who study complexity theory is the existence of efficient algorithms. In the past the existence of polynomial time algorithms has been shown by giving an algorithm and then showing that it can be carried out in polynomial time. Times have changed. Mike Fellow has used Wagoner's conjecture to show that there are polynomial time algorithms for solving certain problems in graph theory. His methods are highly nonconstructive. In fact the existence of these algorithms is probably independent of ZF. So there are now pure existence proofs for computer programs that solve real problems, and no hope

of ever programming them. This might be compared to the situation that existed when Hilbert gave his proof of the Hilbert Basis Theorem which led Gordan to make his famous statement.

Das ist nicht Mathematik. Das ist Theologie.[1]

It seems to me that the interest in constructive mathematics will increase with the growing influence of computers on mathematics. There is already a return of interest in the original problems that led to the abstract structures that most of us have been concerned with. If you want to get in on the ground floor, then the book under review is an excellent starting point.

[1] For an interesting discussion of the controversy that arose over Hilbert's proof of the Basis Theorem see Constance Reid, *Hilbert*, Springer-Verlag, New York, 1970.

TELEGRAPHIC REVIEWS

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T: Textbook	P: Professional Reading	1–4: Semesters
C: Computer Software	L: Undergraduate Library	**: Special Emphasis
S: Supplementary Reading	13: Grade Level	??: Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

General, S. *Succeed With Math: Every Student's Guide to Conquering Math Anxiety.* Sheila Tobias. The College Board (45 Columbus Ave., NY 10023), 1987, xviii + 252 pp, (P). [ISBN: 0-87447-259-8] A popular math-fitness book for students and adults who must learn more mathematics than they think is possible for them. Practical hints on reading and thinking about mathematics, on solving problems, and on uses of mathematics in social science, biology (population genetics), and business. Designed to build self-confidence, this sequel to *Overcoming Math Anxiety* covers many of the mathematical skills required to study freshman-level mathematics. Should be in every college's study skill center. LAS

Mathematics Appreciation, L.** *Mathematical Recreations and Essays, Thirteenth Edition.* W.W. Rouse Ball, H.S.M. Coxeter. Dover, 1987, xvii + 428 pp, $8.95 (P). [ISBN: 0-486-25357-0] Revision of the *12th Edition* (TR, May 1975) published by the University of Toronto Press in 1974. New material features the role of the computer in extending the list of Mersenne primes and in the resolution of the problem of "squaring the square." An addendum on space-filling with "golden" rhombohedra and the connection with quasilattices. Every public or school library should have a copy. No one interested in recreational mathematics can afford to pass up this modestly-priced delightful classic. JK

Mathematics Appreciation, T*(13: 1). *A Mathematics Sampler: Topics for Liberal Arts Students.* William P. Berlinghoff, Kerry E. Grant. Ginn Pr, 1986, 400 pp, (P). [ISBN: 0-536-05450-9] Ambitious book for a freshman course for liberal arts students. Well written. Covers topics such as primes, perfect numbers, geometry, probability and statistics, uncountability of reals, basics of group theory. Looks like an interesting course. MZ

Mathematics Appreciation, T*(13-16: 1). *For All Practical Purposes: Introduction to Contemporary Mathematics.* COMAP (60 Lowell St., Arlington, MA 02174), 1988, xii + 450 pp. [ISBN: 0-7167-1830-8] An innovative textbook for a survey course in mathematics for liberal arts students presenting mathematics through contemporary applications. Written to accompany a telecourse due to be shown on public television beginning in January 1988, this text can also be used alone in a traditional lecture course. Treats management science, statistics, social choice, size and shape, and computers with minimal mathematical prerequisites (just a glimmer of first year high school algebra). Features full-color graphics, interviews with users of mathematics, and many stimulating examples of contemporary mathematics applied to practical problems. JAS

Education, P, L. *Developments in School Mathematics Education Around the World.* Ed: Izaak Wirszup, Robert Streit. NCTM, 1987, x + 725 pp, $20 (P). [ISBN: 0-87353-249-X] Proceedings of a March 1985 international conference sponsored by the University of Chicago School Mathematics Project. One-third of the papers report on mathematics education world wide; others, mostly by Americans, deal with instructional strategies, curriculum design, and technology-supported learning. Includes helpful mini-biographies of all authors. LAS

History, P, L.** *The Rhind Mathematical Papyrus: An Ancient Egyptian Text.* Gay Robins, Charles Shute. British Museum, 1987, 88 pp, (P). [ISBN: 0-7141-0944-4] A complete explication (in modern notation) of the Rhind Papyrus, the most

important "textbook" of Egyptian arithmetic containing rules for adding (unit) fractions followed by problems on solving equations, squaring the circle, calculating volumes and prices. Concludes with full-color plates of the entire papyrus which are linked to the text by schematic diagrams. Contains minimal commentary—just enough to make the text clear without overpowering it. LAS

History, S*(15-17), P, L.** *Studies in the History of Mathematics.* Ed: Esther R. Phillips. Stud. in Math., V. 26. MAA, 1987, 308 pp. [ISBN: 0-88385-128-8] Ten samples of recent research on the history of recent mathematics, principally in the period 1870-1950: from Dedekind's invention of ideals through Killing, Brouwer, Bieberbach, etc., to John von Neumann's marriage of mathematics and computing at the Institute for Advanced Study. Splendid essays suitable as background reading by upperclass mathematics students. LAS

Group Theory, S(18), P. *Lecture Notes in Mathematics-1261: Finite Presentability of S-Arithmetic Groups Compact Presentability of Solvable Groups.* Herbert Abels. Springer-Verlag, 1987, vi + 178 pp, $16.30 (P). [ISBN: 0-387-17975-5]

Group Theory, T(16-17), S, P, L*. *Group Theory.* W.R. Scott. Dover, 1987, 479 pp, $10.95 (P). [ISBN: 0-486-65377-3] Unabridged, corrected republication of a classic 1964 text. Basic definitions and results; abelian groups; p-groups; supersolvable groups; free groups; extensions; permutation groups; representations; products of subgroups; the multiplicative group of a division ring; topics in infinite groups (in view of the Feit-Thompson theorem). Hundreds of exercises. RB

Algebra, P. *Linear Topologies on a Ring: An Overview.* Res. Notes in Math. Ser., V. 159. Jonathan S. Golan. Longman Scientific & Tech (US Distr: Wiley), 1987, 104 pp, $36.95 (P). [ISBN: 0-470-20842-2] Study of the structure of the set of topologizing filters (those inducing a linear topology) of left ideals on a non-commutative ring R. Major theme is the interplay between the order structures and the algebraic structures. RM

Algebra, P. *Lecture Notes in Mathematics-1229: Derivations, Dissipations and Group Actions on C^*-algebras.* Ola Bratteli. Springer-Verlag, 1986, vi + 277 pp, $23.60 (P). [ISBN: 0-387-17199-1] Covers theory of derivations, starts at beginning with bounded derivations, then non-commutative vector fields, and finished with dissipations. MZ

Calculus, P*, L*. *Calculus for a New Century: A Pump, Not a Filter.* Ed: Lynn Arthur Steen. MAA Notes No. 8. MAA, 1988, xiv + 258 pp, $12.50 (P). [ISBN: 0-88385-058-3] Expanded proceedings of an October 1987 colloquium sponsored by the National Academies of Sciences and Engineering intended to stimulate national discussion among scientists, mathematicians, and engineers about revitalizing calculus. Contains plenary papers presented at the colloquium, background papers prepared in advance of the colloquium, surveys of calculus experiments, reports from discussion groups, samples of calculus final examinations, and related readings. Dozens of excellent yet diverse ideas with little focus except discontent with the *status quo.* A valuable resource for would-be calculus reformers; a natural sequel to *Towards a Lean and Lively Calculus* (MAA, 1986, TR, March 1987). JAS

Real Analysis, T(17). *Real Analysis.* William O. Ray. Prentice-Hall, 1988, x + 307 pp. [ISBN: 0-13-762386-0] Aimed at beginning graduate students who have had a course at the level of Goldberg's *Methods of Real Analysis,* this book covers the Lebesgue integral, Banach spaces, integration over abstract spaces, and introductory functional analysis. AWR

Real Analysis, P. *Seminar on Geometric Measure Theory.* R. Hardt, L. Simon. DMV Seminar, Band 7. Birkhauser Boston, 1986, 117 pp, $16.50 (P). [ISBN: 0-8176-1815-5] Ten lectures, five by each author, intended to introduce the main ideas of geometric measure theory to analysts, given over a period of one week at Schloss Mickeln, Düsseldorf. AWR

Differential Equations, S(16). *An Introduction to Kalman Filtering With Applications.* Kenneth S. Miller, Donald M. Leskiw. Robert E Krieger, 1987, vi + 113 pp, $16.50. [ISBN: 0-89874-824-0] Starting with a stochastic differential system (the system model) that predicts the state of a system at time t, we assume that a series of observations are made (the measurement model) which can be used to modify the theoretic model. The Kalman filter is a recursive method that enables one to use all information available up to time t' in order to make a best estimate at time $t^* > t'$. Prerequisite for reading: differential equations, some probability, some mechanics. Readable, attractive text. AWR

Partial Differential Equations, S(17-18), L. *The Mathematical Theory of Huygens' Principle.* Bevan B. Baker, E.T. Copson. Chelsea, 1987, vii + 193 pp, $16.95. [ISBN: 0-8284-0329-5] Reprint with minor changes of the 1950 edition. Huygen's principle is a geometric model of light wave propagation in optics. The book deals with the mathematical theory of Huygens' principle in optics, its application to the theory of diffraction, and the general theory of the solution of the partial differential equations governing light propagation. AM

Partial Differential Equations, T(18). *Nonlinear Partial Differential Equations and Free Boundaries, Volume I: Elliptic Equations.* J.I. Díaz. Res. Notes in Math., V. 106. Pitman, 1985, 323 pp, $44.95 (P). [ISBN: 0-273-08572-7] A readable introduction to the area of nonlinear partial differential equations.

Focuses on two methods of solution, energy methods, and sub-super solution methods. MZ

Partial Differential Equations, P. *Linear Differential Equations of Principal Type.* Yu. V. Egorov. Transl: Dang Prem Kumar. Contemp. Soviet Math. Consultants Bureau, 1986, vii + 301 pp, \$85. [ISBN: 0-306-10992-1] The properties of a differential or pseudo-differential operator of principle type are independent of low-order terms and determined only by the principal symbol. The book presents the topics necessary for the study of differential equations and boundary value problems involving such operators. These include distribution theory, pseudo-differential operators, wave front sets, Fourier integral operators, and results on local solvability of differential equations. AM

Partial Differential Equations, T*(16-17: 1, 2), L. *Introduction to Partial Differential Equations and Hilbert Space Methods, Second Edition.* Karl E. Gustafson. Wiley, 1987, xix + 409 pp, \$46. [ISBN: 0-471-83227-8] Enlarged edition features an expanded treatment of first-order systems, an introduction to computational methods, and some aspects of topical research. At earlier reviewers' behest some straightforward exercises have been added along with four novel, several-page "Pauses" designed to give readers a chance to recollect their thoughts. "Pauses" are intended to be "helpful and hopefully interesting auxiliary material to aid in practice and understanding." Considerably expanded Answers Section. (*First Edition,* TR, February 1981.) JK

Numerical Analysis, P. *Tables for Lagrangian Interpolation Using Chebyshev Points.* Herbert E. Salzer, Norman Levine, Saul Serben. Applied Science, 1984, \$28 (P). [ISBN: 0-915061-00-7] Two tables for use in n-point Lagrangian polynomial interpolation of a function tabulated at the Chebyshev points. Table I lists the Chebyshev points $x_{n,i}$ for every n, and $s_{n,i} = \sin[(2i-1)\pi/2n]$ for odd n, for $n = 2(1)25(5)50(10)100$, to twenty-five significant figures. Table II lists the interpolation coefficients for $n = 20, x = -1(.001)1$, to twenty significant figures. Introductory text includes a discussion of interpolation and use of the tables and references. RH

Numerical Analysis, C(14-17), L. *Nonlinear Parameter Estimation: An Integrated System in BASIC.* MSDOS. John C. Nash, Mary Walker-Smith. Stat.: Textbooks & Mono., V. 82. Dekker, 1987, xiii + 493 pp, \$65. [ISBN: 0-8247-7819-7] A collection of software tools in Basic (on an MSDOS disk inside the cover and reproduced in the text) to estimate nonlinear parameters in mathematical models arising in diverse applications. Focuses on methods minimizing loss functions that operate well on small computers. Text provides commentary on the methods and illustrative applications, but little on algorithms or theory. Extensive bibliography (also available on disk) provides references to literature on the algorithms. LAS

Operator Theory, T(18: 1), S, P. *Spectral Theory of Linear Differential Operators and Comparison Algebras.* H.O. Cordes. Math. Soc. Lect. Note Ser., V. 76. Cambridge U Pr, 1987, ix + 342 pp, \$29.95. [ISBN: 0-521-28443-0] First four chapters supply necessary background in spectral theory and second-order elliptic differential operators followed by six chapters treating various aspects and applications of comparison algebras, defined as $C*$-algebras of singular integral operators with special properties. References, index. JS

Functional Analysis, S(18), P, L. *Theory of Linear Operations.* S. Banach. Transl: F. Jellett. Math. Lib., V. 38. Elsevier Science, 1987, x + 237 pp, \$75.50. [ISBN: 0-444-70184-2] An English translation of a classic work, supplemented by fifty pages on "Some Aspects of the Present Theory" by Pełczyński and Bessaga. Extensive bibliography. JS

Functional Analysis, P*. *Introduction to Various Aspects of Degree Theory in Banach Spaces.* E.H. Rothe. Math. Surv. & Mono., No. 23. AMS, 1986, vi + 242 pp, \$60. [ISBN: 0-8218-1522-9] A nice survey of different topics in degree theory, such as Leray-Schauder degree, Poincaré-Bohl theorem, product theorem, finishing with the linear homotopy theorem. MZ

Functional Analysis, S(18), P. *Semigroups of Linear Operators: An Introduction.* A.C. McBride. Res. Notes in Math. Ser., V. 156. Longman Scientific & Tech (US Distr: Wiley), 1987, 134 pp, \$37.95 (P). [ISBN: 0-470-20824-4] Lecture notes intended to serve as an introduction to the subject, developing the theory as far as existence and uniqueness for Cauchy problems. Applications, references. JS

Functional Analysis, P. *Eigenvalues and s-Numbers.* Albrecht Pietsch. Stud. in Adv. Math., V. 13. Cambridge U Pr, 1987, 360 pp, \$59.50. [ISBN: 0-521-32532-3] Developed on a foundation of abstract operators on Banach spaces, the author identifies as the most important result of the monograph a generalization of Weyl's theorem identifying an operator ideal as being of an optimum eigenvalue type. He then builds a theory of traces, determinants, and other familiar concepts for operators. AWR

Analysis, P. *Generalized Integral Transformations.* A.H. Zemanian. Dover, 1987, xvi + 300 pp, \$7.95 (P). [ISBN: 0-486-65375-7] The subject matter deals with generalizations following the lead of Schwartz's work with generalized functions, applied to such classic integral transformations as Fourier, Laplace, and Hilbert. Assumes familiarity with advanced calculus, some Lebesgue integration, and some complex variable theory. (1968 Interscience edition, TR, June-July 1969.) AWR

Analysis, P. *Operational Calculus Based on the Two-Sided Laplace Integral, Third Edition.* Balth. van der Pol, H. Bremmer. Chelsea, 1987, xiii + 415 pp, \$24.95. [ISBN: 0-8284-0327-9] Intended for application of operational calculus in its modern form to problems in mathematics, physics, and engineering. Originally published in 1950. LC

Analysis, T(17-18: 2). *Measure Theory and Integration.* M.M. Rao. Wiley, 1987, xii + 540 pp, \$59.95. [ISBN: 0-471-82822-X] An attractive new entry in the field of texts for the one-year graduate school real variable course. Author gives explanation as to how his book differs from other standard texts, is thoroughly up-to-date, includes good exercises. By comparison, it is a good value for the price. AWR

Analysis, P. *Lecture Notes in Mathematics-1225: Inverse Problems.* Ed: G. Talenti. Springer-Verlag, 1986, vii + 204 pp, \$19.40 (P). [ISBN: 0-387-17193-2] Texts of lectures and seminars at a CIME session on inverse problems. Focuses on ill-posed and inverse problems, covering inverse eigenvalue problems, Fredholm integral equations of first kind and regularization methods, problems motivated by medical imaging, and others. MZ

Analysis, P. *Jordan Algebras in Analysis, Operatör Theory, and Quantum Mechanics.* Harald Upmeier. CBMS Reg. Conf. Ser. in Math., No. 67. AMS, 1987, vii + 85 pp, \$13 (P). [ISBN: 0-8218-0717-X] Lectures on Jordan algebras and their applications to quantum mechanics, dynamical systems, kernel functions and harmonic analysis, Hua operators, Toeplitz operators, and Toeplitz C^*-algebras. Gives good overview of the topics. MZ

Analysis. *Lecture Notes in Mathematics-1259: Desingularization Strategies for Three-Dimensional Vector Fields.* Felipe Cano Torres. Springer-Verlag, 1987, ix + 194 pp, \$16.30 (P). [ISBN: 0-387-17944-5] Written with the well-initiated expert in mind, the forbidding jargon and notation is complicated by an awkwardness in use of the English language. AWR

Differential Geometry, P. *Lie Groupoids and Lie Algebroids in Differential Geometry.* K. Mackenzie. London Math. Soc. Lect. Note Ser., V. 124. Cambridge U Pr, 1987, xv + 327 pp, \$34.50 (P). [ISBN: 0-521-34882-X] A groupoid is an object which generalizes the algebraic structure of a principal fiber bundle. This book develops the theory of topological groupoids, Lie groupoids, and algebroids, then uses this theory to discuss the cohomology of Lie algebroids and criteria for the existence of connections with a prescribed curvature form. Contains work which is published here for the first time. AM

Differential Geometry, P. *Lecture Notes in Mathematics-1255: Differential Geometry and Differential Equations.* Ed: Gu Chaohao, M. Berger, R.L. Bryant. Springer-Verlag, 1987, xii + 243 pp, \$26.80 (P). [ISBN: 0-387-17849-X] Proceedings of the sixth symposium on differential geometry and differential equations held from June 21 to July 6, 1985 at Fudan University, Shanghai, China. The book contains texts of invited lectures and papers presented at the symposium, many of which are neither published nor due to be published elsewhere. AM

Differential Geometry, T(17-18: 1), P. *Differential Systems and Isometric Embeddings.* Phillip A. Griffiths, Gary R. Jensen. Annals of Math. Stud., No. 114. Princeton U Pr, 1987, xii + 225 pp, \$35; \$15 (P). [ISBN: 0-691-08429-7; 0-691-08430-0] Illustrates how the theory of characteristic varieties for exterior differential systems may be used to study local isometric embeddings. Part of the book's goal is to discuss the notions of involutivity and prolongation through a careful study of the differential exterior system associated to the isometric embedding problem. This study also discusses the essential pointwise algebraic conditions necessary for these embeddings and applies the theory to the problem of embedding Cauchy-Riemann structures. AM

Geometry, P. *Analytical and Geometric Aspects of Hyperbolic Space.* Ed: D.B.A. Epstein. London Math. Soc. Lect. Note Ser., V. 111. Cambridge U Pr, 1987, 323 pp, \$29.95 (P). [ISBN: 0-521-33906-5]; *Low Dimensional Topology and Kleinian Groups.* Ed: D.B.A. Epstein. London Math. Soc. Lect. Note Ser., V. 112. Cambridge U Pr, 1986, 321 pp, \$34.50 (P). [ISBN: 0-521-33905-7] Proceedings of two 1984 symposia at the Universities of Warwick and Durham, UK, which gathered top researchers in hyperbolic geometry, Kleinian groups, and three-dimensional topology. JK

Operations Research, S(16-17), P. *Linear Programming in Infinite-Dimensional Spaces: Theory and Applications.* Edward J. Anderson, Peter Nash. Ser. in Discrete Math. & Optimiz. Wiley, 1987, xi + 172 pp, \$44.95. [ISBN: 0-471-91250-6] Intended to be a comprehensive introduction to linear programming problems posed over infinite-dimensional vector spaces which occur in approximation theory, optimal control theory, etc. Specific examples from a variety of fields are discussed; attention is given to algorithms but not numerical analysis. Directed to graduate students and research workers in mathematical programming and operations research. AWR

Operations Research, P. *Mathematical Modelling Courses.* Ed: J.S. Berry, *et al.* Math. & Its Applic. Halsted Pr, 1987, 281 pp, \$95. [ISBN: 0-470-20836-8] Directed at those who teach, or hope to teach a modelling course at the undergraduate level. Twenty-five articles, all by different authors, tell what has and has not worked in a variety of schools and in a variety of courses. Useful resource, but a high price for 280 pages. AWR

Optimization, P. *Lecture Notes in Control and Information Sciences-93: Stability of Solutions to*

Convex Problems of Optimization. K. Malanowski. Springer-Verlag, 1987, ix + 134 pp, $20.60 (P). [ISBN: 0-387-17589-X] Deals with both global stability (allows the whole domain of values of parameters) and differential stability (only local changes of parameters), but is restricted to problems that are strongly convex, subject to point-wise constraints, satisfying linear independence conditions. Strictly research level. AWR

Optimization, S(16-17), P. *Introduction to Optimization.* Boris T. Polyak. Transl. Ser. in Math. & Engin. Optimization Software, 1987, xxvi + 438 pp, $75. [ISBN: 0-911575-14-6] A comprehensive look at optimization of functions of one or more variables: unconstrained, both differentiable and nondifferentiable; constrained, both linear and nonlinear. Addresses mathematicians and computer scientists, attempting to give mathematical insight without rigorous proof, explanations of how things work without algorithmic details. The introduction itself gives a grand overview. AWR

Control Theory, P. *Lecture Notes in Control and Information Sciences-98: Stochastic Adaptive Control Results and Simulations.* A. Aloneftis. Springer-Verlag, 1987, xii + 120 pp, $20.60 (P). [ISBN: 0-387-18055-9] Study of adaptive control theory (estimators used to produce information that, if known, would enable proper feedback) for single-input time systems. Emphasis on Auto-Regressive Moving Average models with exogenous input for methods of self-timing parameter control. Extensive computer simulations compare theory with practice. RM

Systems Theory, T(16-17: 1), P, L? *Linear Dynamical Systems.* John L. Casti. Math. in Sci. & Eng., V. 135. Academic Pr, 1987, xv + 351 pp, $59.95. [ISBN: 0-12-163451-5] Revision of *Dynamical Systems and Their Applications: Linear Theory,* 1977 (TR, December 1978). Foundations of linear systems treated mathematically but with motivational examples. Discussion of system dynamics, controllability, observability, realization, chapters on theoretical treatment via structure theorems, canonical forms, geometric-algebraic view, and infinite dimensional systems. RM

Probability, P. *Fuzzy Sets Theory and Applications.* Ed: André Jones, Arnold Kaufmann, Hans-Jürgen Zimmermann. NATO ASI Ser. C, V. 177. D Reidel, 1986, xii + 403 pp, $69.50. [ISBN: 90-277-2262-5] Proceedings of the NATO Advanced Study Institute on fuzzy sets theory and their applications. Topics include Radon-Nikodym theorem for fuzzy-set valued measures, decisions with usual values, medical applications with fuzzy sets. MZ

Probability, T(16: 1), S*, P, L.** *Statistical Independence in Probability, Analysis and Number Theory.* Mark Kac. Carus Math. Mono., No. 12. MAA, 1959, xiv + 95 pp, (P). [ISBN: 0-88385-025-7] A paperback reprint of Kac's very popular 1959 Carus monograph that uses techniques of classical analysis to link random numbers, the normal law, statistical independence, prime numbers, kinetic theory, continued fractions, and ergodic phenomena. Kac's style—deep, elegant, and crisp—reveals clues to the profound unity of mathematics. A marvelous source for a senior seminar. LAS

Stochastic Processes, P. *Lectures on Stochastic Flows and Applications.* H. Kunita. Springer-Verlag, 1986, 118 pp, DM 20 (P). [ISBN: 0-387-12878-6] Brownian flows; stochastic differential equations; approximations and limit theorems. Photocopied on grey paper. TAV

Stochastic Processes, P. *The Malliavin Calculus.* Denis R. Bell. Mono. & Surv. in Pure & Appl. Math., V. 34. Longman Scientific & Tech (US Distr: Wiley), 1987, x + 105 pp, $62.95. [ISBN: 0-470-20749-3] An introduction to the subject developed first by Paul Malliavin in 1976. Ideas are drawn from probability theory and functional analysis, creating a tool for developing probabilistic proofs. Gives accounts of approaches of both Stroock and Bismut. AWR

Stochastic Processes, P*. *Extreme Values, Regular Variation, and Point Processes.* Sidney I. Resnick. Appl. Prob., V. 4. Springer-Verlag, 1987, xii + 320 pp, $71. [ISBN: 0-387-96481-9] The topic of extreme values and records is of great current interest. This book lays careful ground work for the point process approach. Well written, with numerous exercises and a substantial bibliography. TAV

Stochastic Processes, P. *Seminar on Stochastic Processes 1985.* Ed: E. Çinlar, K.L. Chung, R.K. Getoor. Prog. in Prob. & Stat., V. 12. Birkhauser Boston, 1986, 324 pp, $29. [ISBN: 0-8176-3331-6] The proceedings of a seminar held in Gainsburg, Florida in March 1985. Seventeen papers on a wide variety of subjects in modern stochastic processes research. TAV

Statistics, P. *Lecture Notes in Statistics-42: Topics in Statistical Information Theory.* S. Kullback, J.C. Keegel, J.H. Kullback. Springer-Verlag, 1987, ix +158 pp, $18.90 (P). [ISBN: 0-387-96512-2] Information theoretic methods are used to study the representation of discrimination information, in particular linear discriminant functionals and informational properties of sub-sigma algebras of the fundamental probability space are used to study stochastic processes and random functions. RM

Statistics, P. *Optimal Paired Comparison Designs for Factorial Experiments.* E.E.M. van Berkum. CWI Tract, V. 31. Math Centrum, 1987, ii + 153 pp, Dfl. 24.20 (P). [ISBN: 90-6196-311-7] In paired comparison experiments, observations are made by presenting pairs of objects to one or more judges. In a round robin, all pairs are presented to

each of the judges. If there are n judges and t objects, then $\binom{n}{t/2}$ comparisons must be made. The aim of this book is to construct designs that are more efficient than the round-robin design. LC

Programming, T(14-15: 1), S, P, L. *Software Engineering with Ada, Second Edition.* Grady Booch. Ser. in Ada and Software Engin. Benjamin/Cummings, 1987, xx + 580 pp, $26.95 (P). [ISBN: 0-8053-0604-8] A textbook on Ada as a tool enabling and promoting sound software engineering principles (careful design, structured coding, software life-cycle management), by a teacher of Department of Defense Ada short courses. Stress on good examples, object-oriented development method. Top-down organization allows brief introduction or intensive study. New edition: new design problems; more complete programs, exercises; chapter summaries. RB

Programming, P. *Advanced Turbo Pascal.* Herbert Schildt. Osborne/McGraw-Hill, 1987, xii + 331 pp, $21.95 (P). [ISBN: 0-07-881283-6] An advanced text on programming that assumes that the reader is quite familiar with the Pascal language in general and the Turbo Pascal compiler in particular. The topics covered include advanced data structures (stacks, queues, linked lists), statistical processes, parsing, and use of the Turbo Pascal toolbox. (1986 McGraw-Hill edition, TR, January 1987.) MS

Languages, T(15: 1). *Program Translation Fundamentals: Methods and Issues.* Peter Calingaert. Software Eng. Ser., V. 22. Computer Science Pr, 1988, xvii + 366 pp, $36.95. [ISBN: 0-88175-096-4] An introductory treatment of programs that translate other programs. This includes such software as compilers, assemblers, macro processors, and interpreters. It covers all the major phases of translation, including syntactic analysis, parsing, code generation, and optimization. It would be appropriate for a one-semester course at the advanced undergraduate or graduate level. MS

Algorithms, T(16-18: 1), P, L. *Probabilistic Analysis of Algorithms: On Computing Methodologies for Computer Algorithms Performance Evaluation.* Micha Hofri. Texts & Mono. in Comput. Sci. Springer-Verlag, 1987, xxv + 240 pp, $25. [ISBN: 0-387-96578-5] Mathematical treatment of algorithm complexity analysis assuming probability distributions on the input data. Emphasis on analytic, probabilistic aspects rather than development of algorithms. Asymptotic analysis, combinatorics used as tools; considers algorithms on permutations (e.g., sorting), communication networks, bin packing heuristics. RM

Computer Systems, S(14-16), P, L. *Real-Time Software Design: A Guide for Microprocessor Systems.* Philip Heller. Birkhauser Boston, 1986, xii + 116 pp, $19.95 (P). [ISBN: 0-8176-3201-8] A brief presentation of design and implementation concepts for small real-time systems, for the novice with programming experience. Introduction to Intel 8086 assembly language, graphical design languages (e.g., structure charts, state diagrams), implementation of data flow diagrams using circular queues and multiple process model, detailed examples (control of intelligent keyboard), optimization maxims. RB

Computer Systems, P. *Lecture Notes in Computer Science-254 & 255: Petri Nets: Central Models and Their Properties.* Ed: W. Brauer, W. Reisig, G. Rozenberg. Springer-Verlag, 1987, (P). *Part I,* x + 480 pp, $34.60 [ISBN: 0-387-17905-4]; *Part II,* x + 516 pp, $37.70. [ISBN: 0-387-17906-2] Petri nets and related elementary net systems are used to model distributed systems and processes. These volumes contain papers based on lectures given at a September 1986 advanced course in Bad Honnef, Germany, including a prologue and an epilogue by Carl Adam Petri. Includes basic theory of nets, diverse applications, and relations to other models of concurrent processes. LAS

Computer Systems, P*. *Managing Programming People: A Personal View.* Philip W. Metzger. Prentice-Hall, 1987, xxv + 157 pp, $20.33. [ISBN: 0-13-551094-5] A memorable essay on the management of people in the context of a programming project's "life cycle;" companion to the author's previous book *Managing a Programming Project.* Humorously written and easily browsed, this brief book offers experienced practical advice and food for thought to managers, particularly to data processing and software engineering professionals. RB

Computer Systems, P. *Computer Capacity Planning: Theory and Practice.* Shui F. Lam, K. Hung Chan. Academic Pr, 1987, xix + 215 pp, $24.95. [ISBN: 0-12-434430-5] The objective of computer capacity planning is to match computer resources to computer workload in a cost-effective manner. This monograph introduces fundamental concepts and evaluates various approaches and techniques in the field, then reports on results of the authors' research study which surveyed computer capacity planning experience in various settings. RB

Computer Graphics, P. *Computer Graphics 1987: Proceedings of CG International '87.* Ed: Tosiyasu L. Kunii. Springer-Verlag, 1987, xi + 490 pp, $113.30. [ISBN: 0-387-70022-6] Proceedings of the fifth international conference held in Japan. Thirty papers give a survey of the state-of-the-art in graphics, including fast algorithms, rendering techniques, hardware, computational geometry, image generation, interfaces, hierarchical modeling. RM

Theory of Computation, P. *Lecture Notes in Computer Science-267: Automata, Languages and Programming.* Ed: Thomas Ottmann. Springer-Verlag, 1987, x + 565 pp, $41.10 (P). [ISBN: 0-387-

18088-5] A collection of 46 papers delivered at the Fourteenth International Conference on Automata, Languages, and Programming held at the University of Karlsruhe in July 1987. The papers cover the following theoretical topics: logic, functional programming, semantics, concurrency, automata theory, complexity theory, and Petri nets. MS

Artificial Intelligence, P. *Pattern Recognition Theory and Applications.* Ed: Pierre A. Devijver, Josef Kittler. NATO ASI Ser. F, V. 30. Springer-Verlag, 1987, xi + 543 pp, $100. [ISBN: 0-387-17700-0] Tutorial and advanced papers from NATO Advanced Study Institute in Belgium, June 1986. Major themes are clustering and relaxation techniques, Markovian and connectionist models, graph theory and geometry, structural methods, hybrid methods and fuzzy sets, knowledge-based techniques, machine vision and image processing. RK

Computer Science, T(18), P. *Stochastic Analysis of Computer Storage.* O.I. Aven, E.G. Coffman, Jr., Y.A. Kogan. Math. & Its Applic. Kluwer Academic, 1987, viii + 254 pp, $89. [ISBN: 90-277-2515-2] A monograph on mathematical analysis of standard problems in computer memory allocation, including fragmentation, page replacement policies, etc., by an American and two Russians. The authors adopt a stochastic statistical model ("sequences of 'arrivals' and 'departures' ") for analysis, shunning combinatorial models. Buffer storage; primary computer storage; paged two-level memories; secondary storage. No exercises. RB

Computer Science, P. *Lecture Notes in Computer Science-248: Networking in Open Systems.* Ed: Günter Müller, Robert P. Blanc. Springer-Verlag, 1987, vi + 441 pp, $30.60 (P). [ISBN: 0-387-17707-8] Proceedings of a one-week seminar, part of IBM Europe Institute at Oberlech, Austria, August 1986. 26 papers on issues concerning open networking standards, some research topics: broad standards issues; emerging end-user applications; programming tools for creating such applications; operating system-communication system interaction; specification, testing. Open standards vs. proprietary systems (such as IBM's). RB

Computer Science, P. *Lecture Notes in Computer Science-246: Graph-Theoretic Concepts in Computer Science.* Ed: Gottfried Tinhofer, Gunther Schmidt. Springer-Verlag, 1987, vii + 306 pp, $25 (P). [ISBN: 0-387-17218-1] Proceedings of the twelfth annual international workshop on the topic, held in a monastary (Bernreid) near Munich, June 1986. Twenty-two papers on specific algorithms, graph grammars, graph manipulation, nets, complexity issues, algorithmic and network considerations, outerplanar graphs, graph isomorphism, parallelism and distributed systems, graphs and geometry, randomness consideration, applications in chemistry. RB

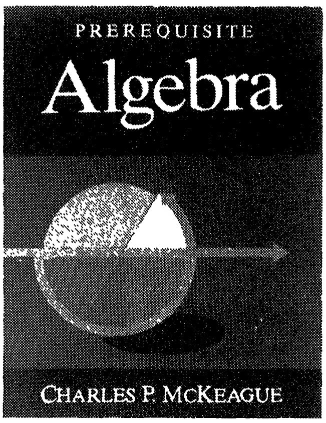
Applications (Engineering), P. *Lecture Notes in Engineering-14: The Boundary Integral Equation Method in Axisymmetric Stress Analysis Problems.* A.A. Bakr. Springer-Verlag, 1986, xi + 213 pp, $18 (P). [ISBN: 0-387-16030-2] On the boundary element method, an efficient and accurate numerical technique and its application to axisymmetric engineering problems. Compares solutions with those obtained analytically or by finite element methods. Ties with problems in elasticity, thermoelasticity, centrifugal, and fracture mechanics. Over 100 references to recent related work. Many practical examples with numerical data and graphs. Authoritatively written. JK

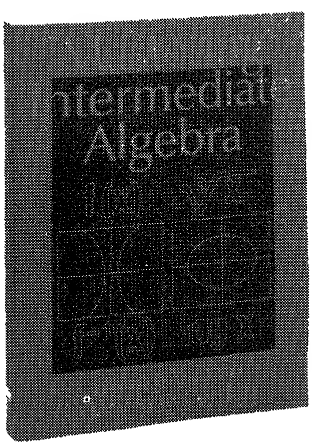
Applications (Management), P, L. *Information Systems Education: Recommendations and Implementation.* Ed: Richard A. Buckingham, *et al.* British Comput. Soc. Mono. in Informatice. Cambridge U Pr, 1987, x + 256 pp, $39.50 (P). [ISBN: 0-521-31234-5] A collection of papers presenting a major British revision of the decade-old IFIP curriculum recommendations for majors in information systems (or what in the United States is often called MIS: management information science). Half the papers discuss curriculum and course outlines; the other half describe teaching methods, social factors, and various concrete case studies. LAS

Applications (Physics), S*(13-15), L. *Space, Time and Gravitation: An Outline of the General Relativity Theory.* Arthur Eddington. Cambridge U Pr, 1987, xii + 218 pp, $14.95 (P). [ISBN: 0-521-33709-7] Reprint of a 1920 classic, one of the most popular science books ever written. A masterpiece of vigorous writing, Eddington's exposition reveals the revolutionary insights of Einstein's theory of relativity without using any technical mathematics or physics. That's one factor in its abiding popularity. LAS

Reviewers

RJA: Richard J. Allen, St. Olaf; DFA: David F. Appleyard, Carleton; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; LC: Laura Chihara, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; SG: Steven Galovich, Carleton; GG: George Gilbert, St. Olaf; BH: Bruce Hanson, St. Olaf; RH: Rhonda Hatcher, St. Olaf; RK: Roger Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; AM: Alan Magnuson, St. Olaf; SM: Steve McKelvey, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; MR: Matthew Richey, St. Olaf; AWR: A. Wayne Roberts, Macalester; MS: Michael Schneider, Macalester; PS: Paul Schoening, St. Olaf; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; LW: Lynne Walling, St. Olaf; FLW: Frank L. Wolf, Carleton; MZ: Michael Zielinski, St. Olaf.


PREREQUISITE
Algebra
CHARLES P. McKEAGUE


Algebra

STATISTICS AND PROBABILITY IN MODERN LIFE, 4/E

Joseph Newmark, *The College of Staten Island, CUNY*

Provides an introduction to the basic concepts of statistics and probability for students in all academic areas. Applications are drawn from areas as diverse as sociology, business, ecology, and medicine.

ISBN 0-03-008367-2 1988 704 pp. (approx.)

TECHNICAL MATHEMATICS, 4/E

Jacqueline Austin and Margarita Isern, *both of Miami-Dade Community College*

Intended for students with varying mathematical backgrounds, this text helps develop problem-solving ability through the use of models and examples featuring a step-by-step approach.

ISBN 0-03-013233-9 1988 736 pp. (approx.)

APPLIED CALCULUS

Dennis Berkey, *Boston University*

Designed for a two-semester course for students of business, biology, and the social sciences.

ISBN 0-03-070733-7 1987 662 pp.

INTERMEDIATE ALGEBRA, 2/E

Jack Barker, James Rogers, and James Van Dyke, *all of Portland Community College*

Intermediate algebra with a focus on manipulative skills and their applications. Taking an adult approach, the authors present concise explanations written with the student in mind.

ISBN 0-03-007068-6 1987 868 pp.

ARITHMETIC, 4/E

Jack Barker, James Rogers, and James Van Dyke, *all of Portland Community College*

A thorough presentation of arithmetic with an emphasis on manipulative skills and their applications. Students are provided insight into the problem-solving procedures and presented with an opportunity to discover the arithmetic concept.

ISBN 0-03-007033-3 1987 648 pp.

BASIC ALGEBRA, 2/E

Jack Barker, James Rogers, and James Van Dyke, *all of Portland Community College*

A work text on elementary algebra with a special approach that is flexible enough to be adaptable for both the classroom lecture format and individualized instruction.

ISBN 0-03-007052-X 1987 750 pp.

MACSYMA

automates symbolic mathematics. And yields enormous improvements in speed, accuracy and modeling power.

If you work with quantitative models in scientific or engineering disciplines, MACSYMA can increase your modeling power. MACSYMA combines symbolic and numeric computation. And enables you to accurately manipulate symbolic expressions in a fraction of the time required manually.

Wide range of capabilities

MACSYMA offers the widest range of capabilities for symbolic computations in applied mathematics of any commercially available program. For example:

Algebra: MACSYMA can manipulate large expressions, expand, simplify and factor expressions, handle matrices and arrays, and solve systems of equations.

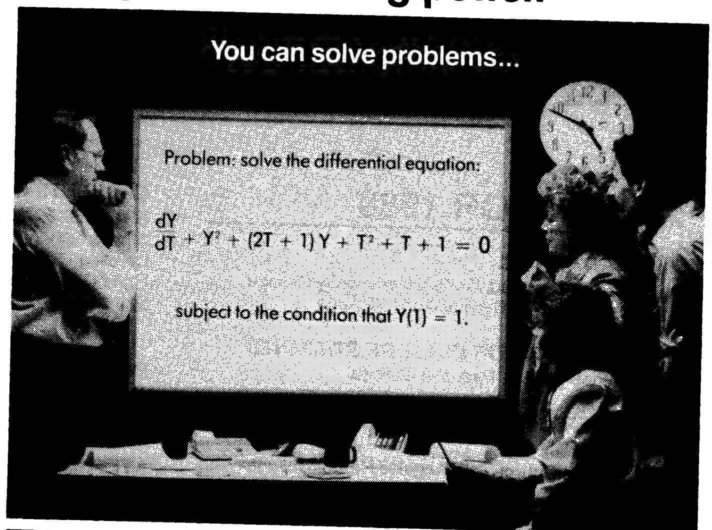
Calculus: MACSYMA can differentiate, perform definite and indefinite integration, take limits, expand functions in Taylor or Laurent series, solve differential equations, and compute Laplace transforms.

Numerical analysis: You can perform numerical analysis using MACSYMA® language, compute with arbitrary precision arithmetic, use a library of numerical analysis routines, and generate FORTRAN code.

Graphics and interfaces: MACSYMA can generate report-quality graphics in 2D or 3D, with perspective, hidden line removal, and captions. MACSYMA also interfaces with the mathematical text processors 'TeX' and 'troff'.

Broad base of applications

Throughout the world thousands of scientists, engineers and mathematicians are using MACSYMA in such diverse applications as aeronautical design, structural engineering, fluid mechanics, acoustics, CAD, electronic and VLSI circuit design, electromagnetic field problems, plasma physics, atomic scattering cross sections, control theory, maximum likelihood estimation, genetic studies, and more.



Symbolically...

```
(C1) DEPENDS(Y,T)$
(C2) DIFF (Y,T) + Y^2 + (2*T + 1)*Y + T^2 + T + 1;
(D2)      dY
          -- + Y^2 + (2 T + 1) Y + T^2 + T + 1
          dT
(C3) SOLN:ODE(D2,Y,T);
(D3)      Y = - %C T %E^T - T - 1
                ----------------
                 %C %E^T - 1
(C4) SOLVE(SUBST([Y = 1, T = 1],D3),%C),NUMER;
(D4)      [%C = 0.5518192]
(C5) SPECIFIC SOLN:SUBST(D4,SOLN);
(D5) Y = - 0.5518192 T %E^T - T - 1
           ------------------------
            0.5518192 %E^T - 1
```

and Numerically.

```
(C6) FORTRAN(D5)$
        Y = - (0.5518192*T*EXP(T) - T - 1)
     1      /(0.5518192*EXP(T) - 1)
```

Available on many computer systems

Current systems include:

- Symbolics 3600™ Series
- VAX & MicroVAX II
- SUN-2 & SUN-3
- Apollo
- Masscomp

Other versions will be following soon.

For an information kit about all the ways MACSYMA can work for you, just call

1-800-MACSYMA.

In Mass., Alaska or Hawaii only, call (617) 621-7770.

Or please write to us at

Computer-Aided Mathematics Group
Dept. M-AM4
Symbolics, Inc.
Eleven Cambridge Center
Cambridge, MA 02142

MACSYMA

The most comprehensive software for mathematical computing.

symbolics™

JOIN MAA TODAY!

Send this application to: MAA Membership Department
1529 18th Street, N.W.
Washington, D.C. 20036

Name ______________________________

Mailing Address ______________________________

______________________________ Zip ______________

Employer/School ______________________________

Position (Rank) ______________________________

Employer's City/State ______________________________

Highest Earned Degree ______________ Year Degree Earned ______________

Institution Awarding Degree ______________________________

Month/Year of Birth ______________________________

Have you been a member of the MAA before? ☐ **Yes** ☐ **No**

Your membership may begin on July 1 or January 1. You may elect to receive any of the listed combinations of the three MAA journals. All members receive Focus, the MAA newsletter.

Please **circle** the appropriate box corresponding to your initial membership period and the selection of journals you wish to receive. (Rates guaranteed for the indicated periods only.) Subscription prices are included with dues (see **).

THE AMERICAN MATHEMATICAL MONTHLY **(M)**
MATHEMATICS MAGAZINE **(G)**
THE COLLEGE MATHEMATICS JOURNAL **(J)**

	M	G	J	M + G	M + J	G + J	M + G + J
Student Membership							
1 year (Jan.-Dec. 1988)	$29	$23	$25	$36	$38	$31	$44
1½ year (July 1987-Dec. 1988)	$41	$32	$35	$51	$54	$44	$63
Regular Membership							
1 year (Jan.-Dec. 1988)	$48	$35	$39	$61	$65	$52	$78
1½ year (July 1987-Dec. 1988)	$68	$49	$54.50	$87	$92.50	$72.50	$111.50

*Student Membership available to high school and undergraduate students and to students regularly enrolled in graduate study at least half time. Enclose letter from school or department official confirming status. Student rates apply to unemployed persons who are seeking employment.

**Annual dues include subscription prices as follows: Reg. Memb. $26 (M), $17 (J), $13 (G), $3 (Focus). Std. Memb. $13 (M), $9 (J), $7 (G), $1 (Focus).

Payment Enclosed $______________ *U.S. Funds Only* (see below)

METHOD OF PAYMENT ☐ Check enclosed ☐ VISA ☐ MasterCard ☐ Bill me

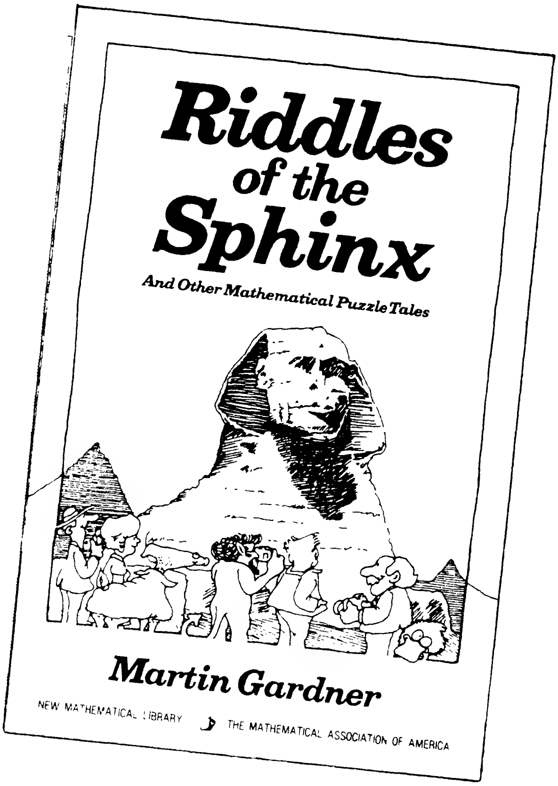
Card Number ______________________________ Expiration Date |_|_|_|_|
(month/year)

Interbank Number |_|_|_|_| *(MasterCard only—located above name on card)*

SIGNATURE ______________________________

Riddles of the Sphinx

and other mathematical puzzle tales.



by Martin Gardner
Volume 32 in the New Mathematical Library
184 pp., Paper. ISBN-0-88285-632-8
List: $14.50 MAA Member: $12.50

Martin Gardner has charmed readers for over fifty years with his delightful books and articles on science. He is best known for the popular column, "*Mathematical Games,*" which appeared in *Scientific American* for twenty-five years. Generations of scientists and mathematicians have been inspired by his writing and the MAA is proud to include his name in its list of authors.

This book was drawn from Gardner's column in *Isaac Asimov's Science Fiction Magazine.* The riddles presented here incorporate the responses of his initial readers, along with additions suggested by the editors of the New Mathematical Library. Each chapter (riddle) poses a problem answered in the First Answers section. The solution in turn raises another problem that is solved in the Second Answers section. This may suggest a third question and in several instances there is a fourth. Gardner draws us from questions to answers always presenting us with new riddles—some as yet unanswered. There are 125 different pieces altogether.

Solving these riddles is not simply a matter of logic and calculation, although these play a role. Luck and inspiration are factors as well, so beginners and experts alike may profitably exercise their wits on Gardner's problems, whose subjects range from geometry to word play to questions relating to physics and geology.

We guarantee that you will solve some of the riddles, be stumped by others, and be amused by almost all of the stories and settings that Gardner has devised to raise these questions.

Order from:
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

THE MATHEMATICAL ASSOCIATION OF AMERICA
1225 Connecticut Avenue, N.W.

THE AMERICAN

MATHEMATICAL MONTHLY

Volume 95, Number 3 March 1988

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Dept. of Mathematics, Stanford University, Stanford, CA 94305 (address effective until May, 1988). Shorter articles about the teaching of mathematics should be sent to STAN WAGON and JOAN HUTCHINSON, Department of Mathematics, Smith College, Northampton, MA 01063. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to ANITA E. SOLOW, Dept. of Mathematics, Grinnell College, Grinnell, IA 50112; in analysis to RICHARD LIBERA, Mathematical Sciences Department, University of Delaware, Newark, DE 19716. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Proposed problems (three copies) and solutions (two copies), both elementary and advanced to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, University of Calgary, Alberta, Canada T2N 1N4.

Reprint Permission: A. B. WILLCOX, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Back Issues: P. and H. BLISS Co., Middletown, CT 06457.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $26 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $29. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $90 per year.

Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Assocation of America, 1529 Eighteenth Street, N.W., Washington, DC 20077-9564.

PRINTED IN THE UNITED STATES OF AMERICA

Angle Trisection, the Heptagon, and the Triskaidecagon

ANDREW M. GLEASON[1], *Harvard University*

Dr. Gleason graduated from Yale in 1942 and then served four years in the Navy. After the war he went to Harvard as a Junior Fellow in the Society of Fellows. Except for an interlude in the Navy from 1950–52, he has been at Harvard ever since. He now holds the Hollis Professorship of Mathematicks and Natural Philosophy, a chair that was endowed in 1727. Although he has no doctor's degree, he says that George Mackey was the equivalent of his dissertation supervisor. He has worked in several areas including topological groups, Banach algebras, finite geometries, and coding theory. He received the Newcomb Cleveland prize of the AAAS in 1952. He is a member of the National Academy of Sciences and is a former president of the American Mathematical Society.

To David Vernon Widder on his 90th birthday

In 1796, Gauss discovered how to construct a regular 17-gon using only ruler and compass. Gauss also showed that regular polygons with 257 or 65537 sides can be constructed. He published these results in his famous *Disquisitiones Arithmeticae* [**3**, section VIII] in 1801. There he gave an analysis of the fields $\mathbf{Q}(\xi)$, where ξ is a complex pth root of unity, p an odd prime, and from this analysis he deduced that a regular n-gon can be constructed if n has the form $2^m p_1 p_2 \cdots p_k$, where $p_1, p_2, \ldots, p_k$ are distinct Fermat primes, that is, odd primes that are one greater than a power of two. He did not give explicit geometric constructions. Gauss also stated [**3**, p. 459] very emphatically that no other regular polygons are constructible, but he never published a proof of this fact. A proof was eventually published by Wantzel [**10**] in 1837. Since no Fermat primes larger than 65537 have been discovered, the list of constructible regular polygons remains as Gauss left it.[2]

If we enlarge our kit of construction tools, other regular polygons may become constructible.[3] For example, given an Archimedean spiral, we can divide any angle into any number of equal parts, and hence draw any regular polygon. Suppose we allow ourselves to trisect angles in addition to the standard ruler and compass constructions, what do we gain? Obviously, we can construct regular polygons with $9, 27, 81, \ldots$ sides, but it is certainly not obvious that we can also make a regular polygon with seven sides. In what follows we shall explore the relation between angle trisections and solving cubic equations and then determine which regular polygons can be constructed with the aid of an angle-trisector. We begin by showing how to draw a regular heptagon.

---

1. The author thanks Ethan Bolker and Persi Diaconis who made valuable suggestions and Lloyd Schoenbach who made the figures.

2. Since $2^{rs} + 1$, with r odd and greater than 1, has the nontrivial factor $2^s + 1$, it is clear that $2^q + 1$ can be prime only if q itself is a power of 2. Thus all Fermat primes have the form $2^{2^k} + 1$. The known Fermat primes $3, 5, 17, 257, 65537$ correspond to $k = 0, 1, 2, 3, 4$. Euler observed that $2^{2^5} + 1 = 641 \times 6700417$, thus disproving Fermat's conjecture that all values of k would yield primes. More recently, it has been shown that $2^{2^k} + 1$ is composite for $6 \leqslant k \leqslant 19$ and many larger values. In many cases the complete factorization has not been found. See [**2**].

3. Both [**1**] and [**6**] contain much information about constructions using extraordinary tools. For an explanation of the standard "ruler and compass" constructions see [**5**].

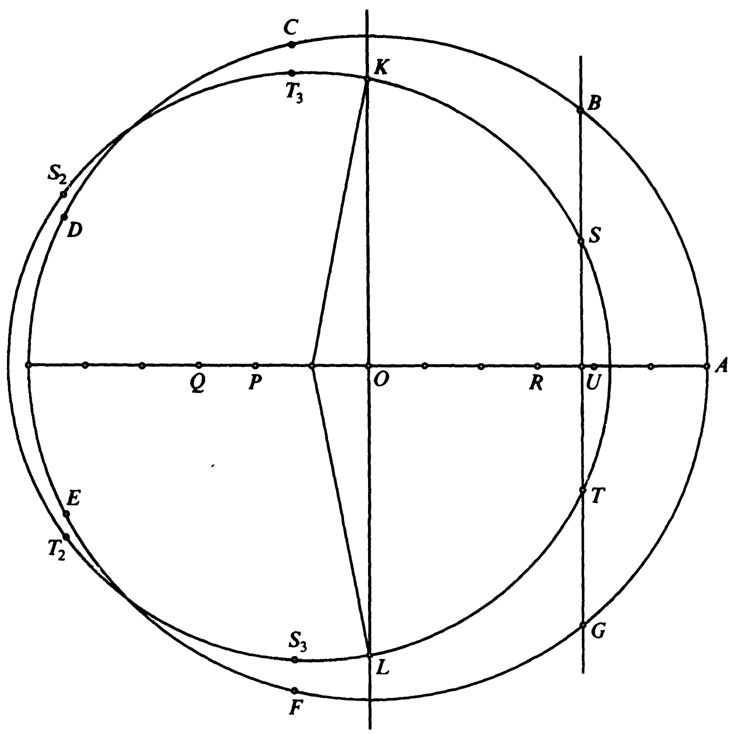


FIG. 1. Construction of a regular heptagon.

Start with a circle $\mathscr{C}$ of radius 6 centered at the origin of a Cartesian coordinate system. Mark $A(6,0)$, $Q(-3,0)$, and $R(3,0)$. Locate $K(0,\sqrt{27})$ and $L(0,-\sqrt{27})$, the vertices of the equilateral triangles with base QR. With center $P(-1,0)$ draw the arc from K to L and trisect it at S and T. The points B and G at which the line ST meets circle $\mathscr{C}$ are vertices of a regular heptagon $ABCDEFG$. The remaining vertices can be found by laying off the arc AB successively around the circle.

Alternatively, we may draw the major arc from K to L and trisect it at S_2 and T_2; then S_2T_2 meets $\mathscr{C}$ at D and E. Finally, if we think of arc KL as going all the way around P, back to K, and then on to L, and trisect this arc at S_3 and T_3, we find C and F.

Proof of the construction. Let U be the point where ST meets the x-axis. It is evident from the construction that $PK=\sqrt{28}$, $\cos\angle APK = 1/\sqrt{28}$, $PU = \sqrt{28}\cos\angle APS = \sqrt{28}\cos(1/3)\angle APK$. The construction is correct if $OU = 6\cos 2\pi/7$; or equivalently, if $PU = 1 + 6\cos 2\pi/7$. Thus we need only establish the identity $\sqrt{28}\cos((1/3)\operatorname{arc}\cos(1/\sqrt{28})) = 1 + 6\cos(2\pi/7)$. To do so, let $\xi = \cos 2\pi/7 + i\sin 2\pi/7$ be the principal seventh root of 1. Put $\eta = \xi + \xi^{-1} = 2\cos 2\pi/7$. Then $\eta^3 + \eta^2 - 2\eta - 1 = \xi^3 + \xi^2 + \xi + 1 + \xi^{-1} + \xi^{-2} + \xi^{-3} = 0$.

Thus η is a root of the equation

$$X^3 + X^2 - 2X - 1 = 0. \tag{1}$$

Setting $X = (Y - 1)/3$, we find that $1 + 6\cos 2\pi/7$ $(= 1 + 3\eta)$ is a root of

$$Y^3 - 21Y - 7 = 0. \tag{2}$$

The substitution $Y = \sqrt{28}\cos\theta$ reduces this to

$$7\sqrt{28}\,(4\cos^3\theta - 3\cos\theta) = 7;$$

whence

$$\cos 3\theta = \frac{1}{\sqrt{28}}.$$

This leads to six determinations of θ modulo 2π, which pair off to give three roots of (2). We leave it to the reader to check that $1 + 3\eta$ corresponds to the choice $\theta = 1/3 \arccos 1/\sqrt{28}$. The other roots of (2) are, of course, $1 + 6\cos 4\pi/7$ and $1 + 6\cos 6\pi/7$; these correspond to different determinations of θ and lead to the alternative constructions of the points C, D, E, and F mentioned above.

Plemelj [6] gave a different construction of the regular heptagon $ABCDEFG$, given A and the center O:

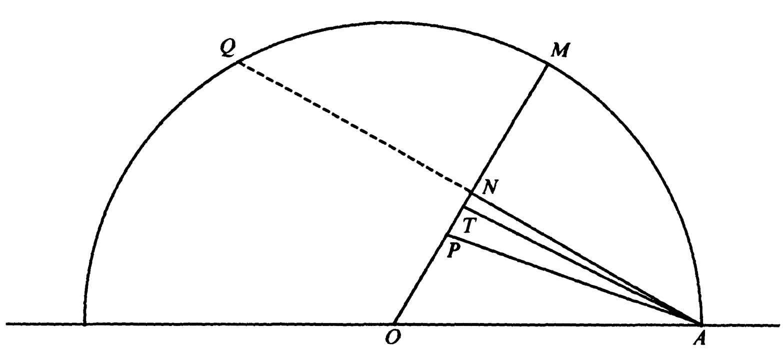


FIG. 2. Plemelj's construction.

Draw the circle with center O passing through A and on it find M so that $AM = OA$. Bisect and trisect OM at N and P, respectively, and find T on NP so that $\angle NAT = \frac{1}{3}\angle NAP$. Then AT equals the side of the required heptagon, which can easily be completed by laying off this segment around the circle.

Plemelj goes on to note that because the angle to be trisected is so small, the simplest approximate trisection will introduce an error too small to be noticeable in any practical case. If we take T to be one-third of the way from N to P, then AT will be too long by less than one part in 20000. Even if we take T at N, AT will be too short by only about one part in 400. Since AN is half of AQ, the side of an inscribed equilateral triangle, we have the rule: The side of the inscribed regular heptagon is (approximately) half of the side of the inscribed equilateral triangle. According to Tropfke **[8]**, this approximation was used by the tenth-century Arabian mathematician Abul Wafa Mohamed, and was also known to Heron of Alexandria, and perhaps even to earlier mathematicians.

To validate Plemelj's construction, we must prove that $AT = OA(2\sin\pi/7)$. Since $2\cos 2\pi/7 = 2 - (2\sin\pi/7)^2$, it follows from (1) that $2\sin\pi/7$ is a root of the equation

$$(2 - X^2)^3 + (2 - X^2)^2 - 2(2 - X^2) - 1 = 0.$$

The other roots are $-2\sin\pi/7$, $\pm 2\sin 2\pi/7$, and $\pm 2\sin 3\pi/7$.
This equation factors:

$$\left(X^3 + \sqrt{7}(X^2 - 1)\right)\left(X^3 - \sqrt{7}(X^2 - 1)\right) = 0.$$

The zeros of the first factor are $2\sin\pi/7$, $-2\sin 2\pi/7$, and $-2\sin 3\pi/7$. If we write the corresponding equation in the form

$$\left(\frac{1}{X}\right)^3 - \frac{1}{X} = \frac{1}{\sqrt{7}}, \tag{3}$$

and make the standard substitution $1/X = 2\sqrt{1/3}\cos\psi$, then (3) becomes $\cos 3\psi = \sqrt{27/28}$. The desired root corresponds to the choice

$$\psi = \alpha = \tfrac{1}{3}\operatorname{arc\,cos}\sqrt{27/28} = \tfrac{1}{3}\operatorname{arc\,tan} 1/3\sqrt{3},$$

and we have finally

$$2\sin\frac{\pi}{7}\cos\alpha = \frac{1}{2}\sqrt{3}.$$

From FIGURE 2 we see that $\angle NAP = \operatorname{arc\,tan} 1/3\sqrt{3}$, so $\angle NAT = \alpha$. Hence we have $AT\cos\alpha = AN = (1/2)\sqrt{3}\,OA$. Comparing these equations, we see that $AT = OA(2\sin\pi/7)$, as required.

Note that $\angle OKP$ in FIGURE 1 is the same as $\angle NAP$ in FIGURE 2. Thus $\angle NAP$ (FIGURE 2) is the complement of $\angle OPK$ (FIGURE 1), so $\angle OPS = \pi/6 - \alpha$. Thus the trisections required in the two constructions are equivalent in the sense that either trisector can readily be constructed from the other by ruler and compass.

It is interesting to pursue the other two roots of (3). They are given, of course, by changing the determination of ψ from α to $\alpha + 2\pi/3$ and to $\alpha - 2\pi/3$. We have

$$\left(-2\sin\frac{2\pi}{7}\right)\cos\left(\alpha + \frac{2\pi}{3}\right) = \left(-2\sin\frac{3\pi}{7}\right)\cos\left(\alpha - \frac{2\pi}{3}\right) = \frac{1}{2}\sqrt{3}.$$

Absorbing the unwanted minus signs into the cosines, we find

$$2\sin\frac{2\pi}{7}\sin\left(\frac{\pi}{6} + \alpha\right) = 2\sin\frac{3\pi}{7}\sin\left(\frac{\pi}{6} - \alpha\right) = \frac{1}{2}\sqrt{3}.$$

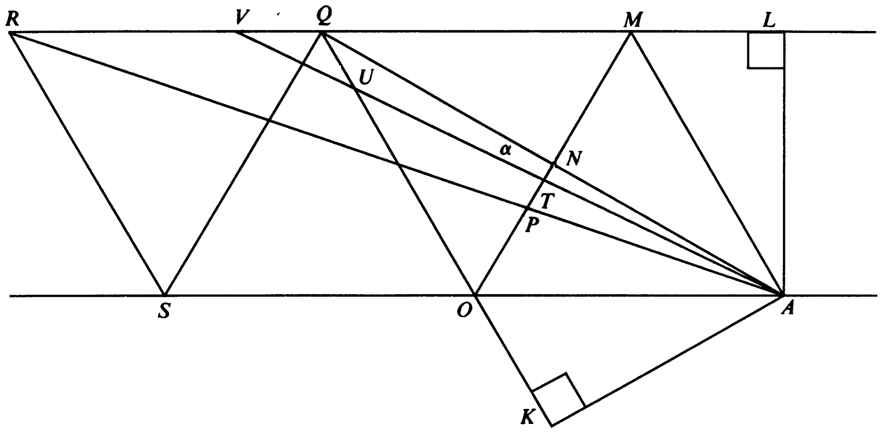


FIG. 3. Plemelj's construction extended. O, A, M, Q, R, S are points of an equilateral triangular lattice. Line AV is chosen so that $\angle QAV = 1/3\angle QAR$. Then AT, AU, and AV equal, respectively, the side, short diagonal, and long diagonal of the regular heptagon inscribed in a circle of radius OA.

In FIGURE 3 we have $AU\sin(\pi/6 + \alpha) = AK = (1/2)\sqrt{3}\,OA$ and $AV\sin(\pi/6 - \alpha) = AL = (1/2)\sqrt{3}\,OA$. Hence $AU = OA(2\sin 2\pi/7)$ and $AV = OA(2\sin 3\pi/7)$. This shows that AU and AV are equal, respectively, to the short and long diagonals of the heptagon.

Now let us consider the theory underlying these constructions. We begin by reviewing the solution of the cubic equation with real coefficients. The term in X^2 can always be removed by translating the roots and, to avoid fractions later on, we write the equation in the form

$$X^3 - 3pX + 2q = 0. \tag{4}$$

Since this equation has odd degree, it must have at least one real root. The nature of the other roots can be found by considering the quantity[4]

$$D = q^2 - p^3.$$

If $D > 0$, there are two conjugate complex roots and one real root; if $D < 0$, there are three distinct real roots; and if $D = 0$, there is one double root, q/p, and a simple root, $-2q/p$ (unless $p = q = 0$, in which case 0 is obviously a triple root). These facts are easily deduced by examining the critical points of the polynomial function in (4). Note that D cannot be negative unless p is positive.

For positive D, Cardano's formula gives the unique real root

$$\sqrt[3]{-q + \sqrt{D}} + \sqrt[3]{-q - \sqrt{D}}\,.$$

This root could be constructed from p and q by ruler and compass and the extraction of one cube root. (The second cube root can be constructed from the first by ruler and compass, since the product of the two cube roots is p.)

When $D < 0$, we have what is known as the *casus irreducibilis* and a seeming paradox: although all the roots are real, they cannot be found by radicals without leaving the real domain.[5] Cardano's formula remains valid, but it involves the cube roots of complex numbers.

To find the cube root of a complex number c, we must in general find the cube root of the real number $|c|$ and trisect the polar angle of c. In this special situation, however, only the trisection requires special tools since the absolute value of either cubic radicand is $p^{3/2}$, for which the cube root can easily be constructed.

When we know that all the roots are real, we can jump directly to the trisection problem by setting $X = 2\sqrt{p}\cos\theta$ in (4), which converts this equation into $\cos 3\theta = -qp^{-3/2}$. Note that the hypothesis $D < 0$ guarantees that $|qp^{-3/2}| < 1$, so there is a constructible angle to trisect. As before, the six determinations of θ modulo 2π lead to the three desired roots. After one value of θ has been obtained, the others can easily be found by adding and subtracting $2\pi/3$; hence one angle-trisection will suffice to find all three roots.

Conversely, any cubic equation that can be solved by an angle trisection must have all its roots real, since the method will produce three roots if it produces any. Thus we have the fundamental result:

THEOREM 1. *A real cubic equation can be solved geometrically using ruler, compass, and angle-trisector if and only if its roots are all real.*

In particular, an angle trisector will not help us to duplicate the cube, because the equation to be solved, $X^3 = 2$, has only one real root.

---

4. This discussion is usually given in terms of the discriminant of the equation (i.e., the square of the product of all the differences of the roots). The discriminant of (4) is not D, but $-108D$; consequently positive D corresponds to negative discriminant and vice versa.

5. See [9, p. 180] for a proof of this fact. See [4] for an extended discussion of the solution of equations by real radicals.

We are now in position to describe precisely what can be constructed with ruler, compass, and angle-trisector.

Associated with any geometric figure (i.e., a finite collection of points, lines, and circles) in a Cartesian coordinate plane[6] is a certain subfield of the real numbers, namely, the field generated by the coordinates of all the points and the coefficients of the equations of all the lines and circles when written in the standard forms $y = mx + b$ (or $x = a$) and $x^2 + y^2 = ax + by + c$. Suppose the data of a construction problem[7] are associated with the field F_0 and the figure to be constructed is associated with the field G. The construction can be carried out with ruler and compass alone if and only if there is a tower of fields $F_0 \subseteq F_1 \subseteq F_2 \subseteq \cdots \subseteq F_k$ such that $G \subseteq F_k$ and each F_i $(i = 1, \dots, k)$ is obtained from F_{i-1} by adjoining the square root of some positive element of F_{i-1}. (See [9, p. 183ff.] for the proof of this standard result.) The intermediate fields F_i correspond to the original figure augmented by the successive points, lines, and circles used in the construction. When we allow the use of the trisector, the result is the same except that we now also allow ourselves to build F_i from F_{i-1} by adjoining a root of a cubic polynomial having coefficients in F_{i-1} and all real roots. The proof of this new theorem is virtually identical to the proof of the standard theorem, once Theorem 1 has been established.

It is convenient to describe the above situation by saying that the field F_k can be constructed from F_0.

For regular polygons we have the following theorem:

THEOREM 2. *A regular polygon of n sides can be constructed by ruler, compass, and angle-trisector if and only if the prime factorization of n is* $2^r 3^s p_1 p_2 \cdots p_k$, *where* $p_1, p_2, \dots, p_k$ *are distinct primes* (> 3) *each of the form*[8] $2^t 3^u + 1$. (*We include the possibility* $k = 0$; *i.e.*, $n = 2^r 3^s$.)

The proof uses the following lemma.

LEMMA. *Suppose K is a real field and L is a normal extension of K of degree* 3. *Then L can be constructed from K by ruler, compass, and angle-trisector.*

Proof. We know that $L = K[\beta]$, where β is a zero of some irreducible cubic polynomial $p(X)$ with coefficients in K. Since L is a normal extension of K, all of

6. We assume that the points $(0,0)$ and $(1,0)$ are points of the figure.

7. If a coordinate system is not given in the plane of the data, first construct one, choosing two given points as $(0,0)$ and $(1,0)$. The data must include two points, or at least two points of intersection must be immediately at hand, else no constructions can be carried out. (We do not accept the instruction "choose a point at random" because that complicates the analysis of constructions considerably. See [5].)

8. Altogether there are 41 primes of this form less than one million; namely, 2, 3, 5, 7, 13, 17, 19, 37, 73, 97, 109, 163, 193, 257, 433, 487, 577, 769, 1153, 1297, 1459, 2593, 2917, 3457, 3889, 10369, 12289, 17497, 18433, 39367, 52489, 65537, 139969, 147457, 209953, 331777, 472393, 629857, 746497, 839809, and 995329. It is reasonable to conjecture that there are infinitely many; probably about $9t$ of them less than 10 .

the zeros of $p(X)$ lie in L, and any one of them generates L. But one of the zeros, say γ, is real and $L = K[\gamma]$. Thus, L is a real field, and the zeros of $p(X)$ are all real. Hence the lemma follows from Theorem 1.

Proof of Theorem 2. Since the proof is very similar to the well known proof for ruler and compass alone, we give only the most important steps.

Suppose n is an integer, at least 3. Let $\xi = e^{2\pi i/n} = \cos 2\pi/n + i \sin 2\pi/n$, and $\eta = \xi + \xi^{-1} = 2\cos 2\pi/n$. The Galois group over $\mathbb{Q}$ of the cyclotomic field $\mathbb{Q}(\xi)$ is abelian with $\varphi(n)$ elements, where φ is Euler's phi function. Consequently, every field between $\mathbb{Q}$ and $\mathbb{Q}(\xi)$ is normal over $\mathbb{Q}$ with abelian Galois group. In particular, the real field $\mathbb{Q}(\eta)$ is normal over $\mathbb{Q}$. Since ξ has degree 2 over $\mathbb{Q}(\eta)$, the degree of $\mathbb{Q}(\eta)$ over $\mathbb{Q}$ is $(1/2)\varphi(n)$.

Now suppose n has the form stated in Theorem 2. Then $\varphi(n) = 2^v 3^w$ for some integers v and w, so the Galois group of $\mathbb{Q}(\eta)$ has $2^{v-1}3^w$ elements. This group will, therefore, have a composition series of length $v + w - 1$ with all quotients isomorphic either to $\mathbb{Z}_2$ or $\mathbb{Z}_3$. Correspondingly, there is a tower

$$F_0 = \mathbb{Q} \subseteq F_1 \subseteq \cdots \subseteq F_{v+w-1} = \mathbb{Q}(\eta)$$

of real fields, each normal over its predecessor of degree 2 or 3. Applying the lemma, we see that the field $\mathbb{Q}(\eta)$ can be constructed using ruler, compass, and angle-trisector. This means we can construct a segment of length $\cos 2\pi/n$, and from this we can easily construct a regular n-gon. Looking back, we see that we will have to use the angle-trisector exactly w times.

Conversely, suppose a regular n-gon can be constructed with ruler, compass and trisector. Then η can be constructed, so it must lie in some field of degree $2^a 3^b$ over $\mathbb{Q}$. Hence η itself has degree $2^c 3^d$, and $\varphi(n) = 2^{c+1}3^d$. But this implies that n has the form given in the theorem.

As an application of our theory, consider the next new prime, $13 = 2^2 \cdot 3 + 1$. Theorem 2 tells us that the regular triskaidecagon can be constructed using one angle trisection. There are many ways to proceed; none seem geometrically perspicuous.

The numbers $2\cos 2\pi k/13$, $k = 1, \ldots, 6$ are the zeros of the polynomial

$$X^6 + X^5 - 5X^4 - 4X^3 + 6X^2 + 3X - 1,$$

which factors over the field $\mathbb{Q}(\sqrt{13})$ to

$$\left(X^3 - X - 1 + \lambda(X^2 - 1)\right)\left(X^3 - X - 1 + \bar{\lambda}(X^2 - 1)\right),$$

where $\lambda = (1 - \sqrt{13})/2$ and $\bar{\lambda} = (1 + \sqrt{13})/2$. The first factor has the zero $2\cos 2\pi/13$. Then $2\lambda + 12\cos 2\pi/13$ is a zero of a cubic polynomial having no quadratic term which we can find as above. After considerable computation we obtain

$$12\cos\frac{2\pi}{13} = \sqrt{13} - 1 + \sqrt{104 - 8\sqrt{13}}\cos\frac{1}{3}\arctan\frac{\sqrt{3}(\sqrt{13} + 1)}{7 - \sqrt{13}},$$

which leads to the following construction, quite similar to the first one for the heptagon:

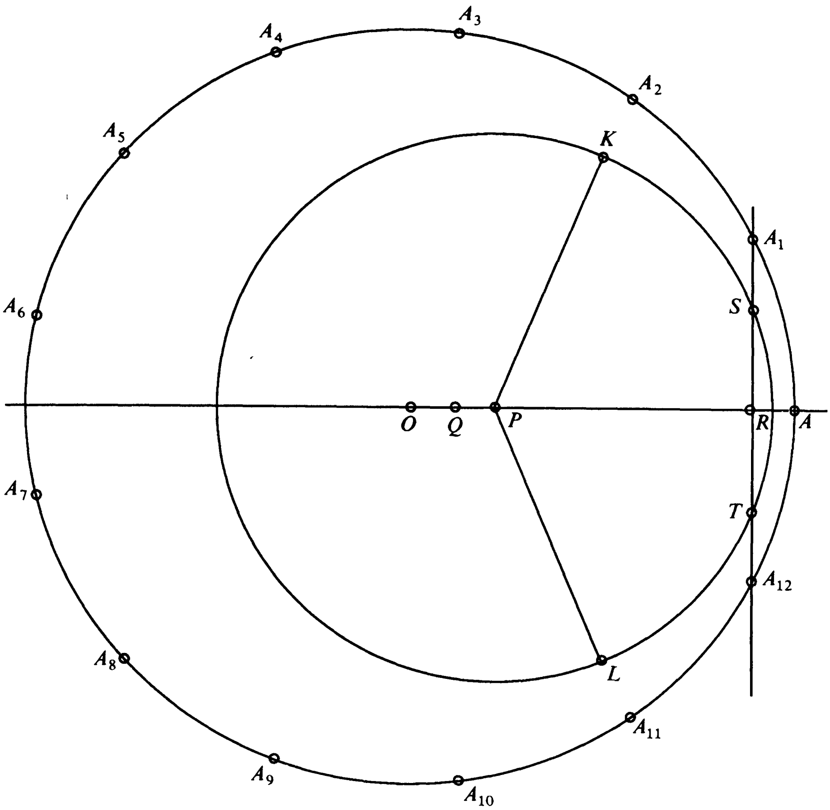


FIG. 4. Construction of a regular triskaidecagon.

Let $\mathscr{D}$ be the circle of radius 12 with center at the origin. Mark $A(12,0)$, $P(\sqrt{13}-1,0)$, $Q(5-\sqrt{13},0)$, and $R(7+\sqrt{13},0)$. Locate $K(6,\sqrt{3}(\sqrt{13}+1))$ and $L(6,-\sqrt{3}(\sqrt{13}+1))$, the vertices of the equilateral triangles having base QR. With center P draw the arc KL and trisect it at S and T. The line ST meets $\mathscr{D}$ at A_1 and A_{12}, vertices of the regular triskaidecagon $AA_1A_2\cdots A_{12}$.

Note that a segment of length $\sqrt{13}$ is easily found as the hypotenuse of a right triangle with sides 2 and 3.

The figure suggests two approximate constructions. The point R appears to be very near the line ST. Indeed, the line through R perpendicular to the x-axis meets $\mathscr{D}$ at a point about twelve minutes of arc from A_1. Even closer, the line PK meets $\mathscr{D}$ at a point within three minutes of A_2.

The next new prime, 19, requires two trisections. The details are left to the reader!

Possible generalizations of Theorem 2 immediately suggest themselves. To construct a regular 11-gon, we must solve the fifth-degree equation having the root $2\cos 2\pi/11$. Can this be done by quinsecting some angle? Gauss answered all such questions. Here is his final statement on the subject [3, p. 450][9]:

> As a result the division of the whole circle into n [a prime] parts requires, *first*, the division of the whole circle into $n-1$ parts; *second*, the division into $n-1$ parts of another arc which can be constructed as soon as the first division is accomplished; *third*, the extraction of one square root, and it can be shown that this is always $\sqrt{n}$.

It follows easily from this that a regular n-gon (n need no longer be prime) can be constructed if, in addition to ruler and compass, equipment is available to p-sect any angle for every prime p that divides $\varphi(n)$. Thus, ruler, compass, and angle quinsector will suffice to construct a regular 11-gon, 41-gon, or 101-gon.

REFERENCES

1. L. Bieberbach, Theorie der Geometrischen Konstruktionen, Birkhauser, Basel, 1952.
2. J. Brillhart, D. H. Lehmer, J. L. Selfridge, B. Tuckerman, and S. S. Wagstaff, Factorizations of $b^n \pm 1$, *Contemporary Mathematics*, 22 (1983) American Mathematical Society, Providence, R.I.
3. C. F. Gauss, Disquisitiones Arithmeticae, Leipzig, 1801, translated from the Latin by A. A. Clarke, Yale University Press, New Haven, CT, 1965.
4. I. M. Isaacs, Solution of polynomials by real radicals, *Amer. Math. Monthly*, 92 (1985) 571–575.
5. N. D. Kazarinoff, Ruler and the Round, vol. 15, Complementary series in mathematics, Prindle, Weber and Schmidt, Boston, 1970.
6. H. Lebesgue, Leçons sur les constructions géométriques, Gauthier-Villars, Paris, 1950.
7. J. Plemelj, Die Siebenteilung des Kreises, *Monatshefte fur Math. und Phys.*, 23 (1912) 309–311.
8. J. Tropfke, Geschichte der Elementar-Mathematik, Band IV, de Gruyter, Berlin, 1940, p. 259.
9. B. L. van der Waerden, Modern Algebra (trans. from the German by Fred Blum), Frederick Ungar Publishing Co., New York, 1949.
10. P. L. Wantzel, *Journal de Mathématiques pures et appliquées*, 2 (1837) 366–372.

9. Quoted by permission of Yale University Press.

New Invariants in the Theory of Knots

LOUIS H. KAUFFMAN

LOUIS H. KAUFFMAN received his Bachelor of Arts degree at MIT and his Ph.D. degree from Princeton University in 1972. He has been at the University of Illinois at Chicago since 1971, where he is now Professor of Mathematics.

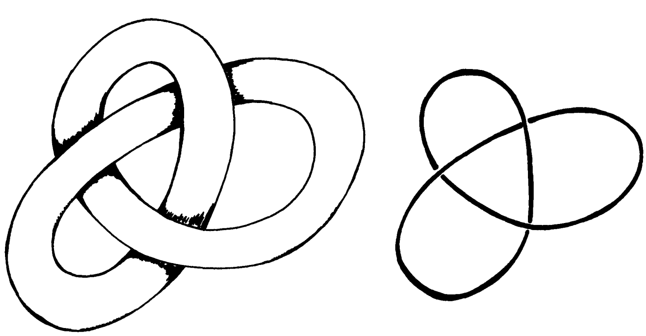
His interest in the theory of knots began at Princeton University through conversations with William Browder (his thesis advisor), Ralph Fox, fellow students, and teachers. He has worked on the relationship of knot theory with singularities of algebraic varieties, manifolds of many dimensions, circuit theory, logic, formal mathematics and physics.

He is grateful to the knots for being such a fertile place in the meeting of ideas and the people who come along with these ideas and give them life.

I. Introduction

In this article I will concentrate on a diagrammatic approach to invariants of knots. I will talk about connections with graph theory, physics, and other topics. In the process I shall construct the Jones polynomial and its associated algebra. I will also discuss generalizations of the Jones polynomial due to myself and others. [Jones introduced his polynomial in 1984. Almost immediately, Hoste, Ocneanu, Millett, Lickorish, Freyd, Yetter, Przytcki, and Traczwyk had a significant generalization. Shortly, yet another invariant was crafted by Brandt, Lickorish, Millett, and Ho. I generalized this one, and in the process found new approaches to the original Jones polynomial.] We'll also explain how proofs of some old conjectures about alternating knots emerge from this work (due to myself, Murasugi, and Thistlethwaite). Many people have helped in this resurgence of the theory of knots. This article is dedicated to all of them.

Let's begin by thinking about how one might go about making a theory of knots and links in three-dimensional space. The typical example of a knot is illustrated in FIGURE 1.



a (realistic) *b* (schematic)

FIG. 1.

Actually, FIGURE 1 has two illustrations of a trefoil knot T. In the first (a), the trefoil is depicted realistically as a physical tube with thickness and shading in the three-dimensional space. This picture reminds you that the trefoil might be made of rope or rubber—and that such a model would exhibit thickness, tension, friction, and other physical properties. In the second illustration (b), there is a schematic representation consisting of three continuous planar segments meeting at *crossings*. The crossings have local forms as shown in FIGURE 2.

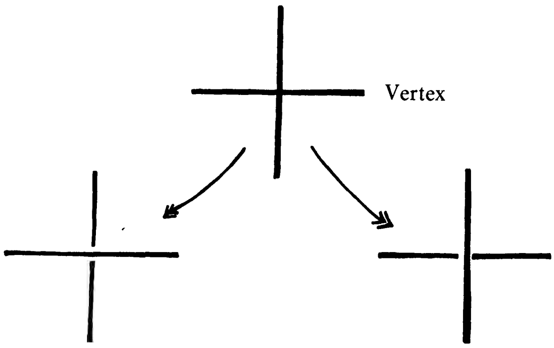


FIG. 2. Two forms of crossing associated to a vertex.

A schematic diagram of this type is a sufficient pattern to allow reconstruction of the knot or link from rope or string. It also encodes key topological properties, and allows the construction of a diagrammatic theory.

Thus, I shall regard a knot or link as extra structure (via crossing choices) on a (locally) 4-valent planar graph. Each vertex of such a graph has the form seen in FIGURE 2, and we shall call such a graph a *universe*. Thus in FIGURE 3 you see the trefoil and its universe.

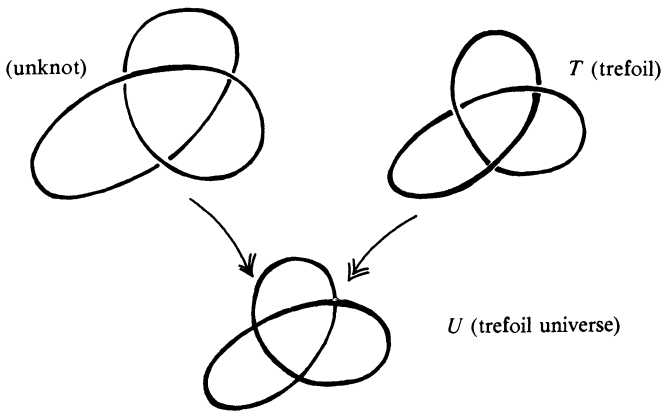


FIG. 3.

As FIGURE 3 indicates, you may regard the universe as the shadow, under projection to the plane, of the overlying knot or link. In general a universe of n vertices can be the projection of 2^n corresponding knots/links. Many of these will be unknotted or unlinked (see FIGURE 3).

But we now need a definition of equivalence so that the words unknotted and unlinked make sense. This equivalence is essentially generated by three fundamental types of diagram moves (the Reidemeister moves). See FIGURE 4. I have designated the Reidemeister moves as type I (add or remove a curl), type II (remove or add two consecutive under (over) crossings), and type III (triangle move). Reidemeister proved in the 1920s that these three moves (in conjunction with planar topological equivalences of the underlying universes) are sufficient to generate spatial isotopy. In other words, Reidemeister proved that two knots (links) in space can be deformed into each other (ambient isotopy) if and only if their diagrams can be transformed into one another by planar isotopy and the three moves (see [**R**]).

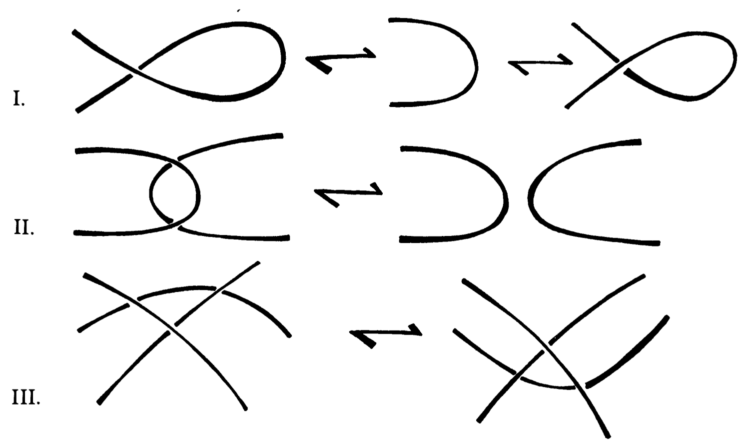


FIG. 4. Reidemeister moves.

By a planar isotopy I just mean a motion of the diagram in the plane that preserves the graphical structure of the underlying universe. See FIGURE 5.

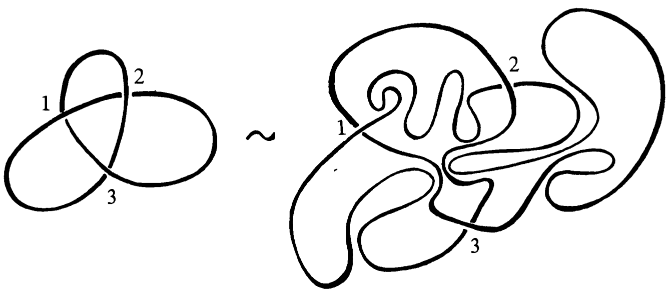


FIG. 5. Planar isotopy.

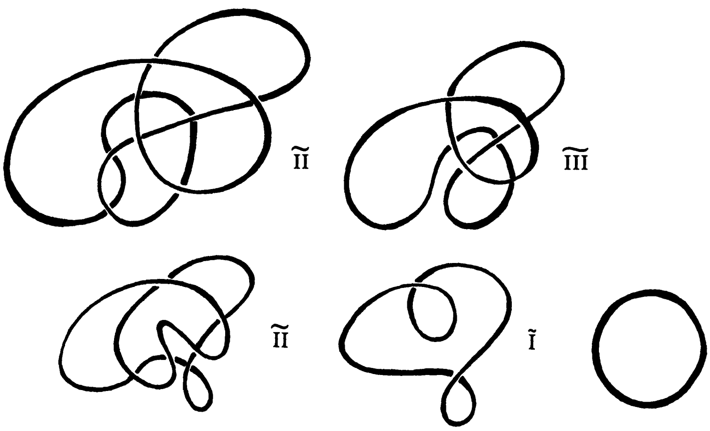


FIG. 6. Ambient isotopy.

In FIGURE 6 an ambient isotopy to an unknotted circle is shown. FIGURE 7 illustrates an ambient isotopy between the figure eight knot E and its mirror image E^*. A knot (link) is said to be ambient isotopic (equivalent) to another if there is a sequence of Reidemeister moves and planar equivalences between them. We write $\sim$ for equivalence. Thus $E \sim E^*$. (Note that the last two steps in the FIGURE 7 deformation are planar equivalences.)

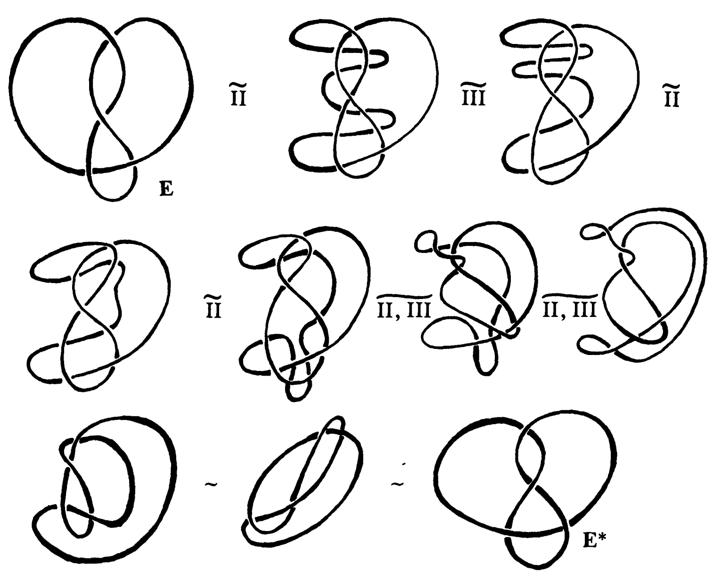
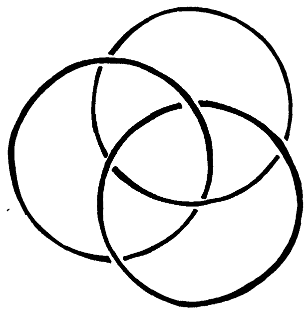


FIG. 7.

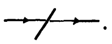
The next thing one wants in a theory of knots are *methods for distinguishing inequivalent knots and links*. For example, we know that there is no equivalence between  and  : for the *number of components remains invariant under* $\sim$. Note that you can determine the number of components even from a complicated diagram by choosing a point on some arc of the diagram and then taking a walk along the diagram—crossing crossings when you come to them. Each component is a complete cycle obtained in this way. (See FIGURE 8).



3 components

FIG. 8. Counting components.

Note indeed that it is a consequence of the Reidemeister moves that the number of components is unchanged by equivalence. Thus the component count is our first invariant of knots and links. By itself it is, however, not very strong. In particular,  and  each have two components but are, in fact, inequivalent. The next invariant is the *linking number*. It gives a measure of how two curves wrap around each other. To define it we need notions of *orientation* and *sign*.

A link is said to be *oriented* if each of its components is assigned a direction indicated by arrow(s) on its arcs. The arrows are consistently arranged in the form . Oriented crossings are given *signs* of ± 1 as shown in FIGURE 9.

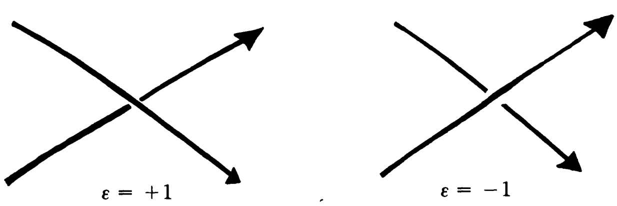


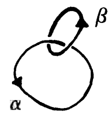
FIG. 9. Crossing signs.

Given a link of two components α and β, let $\alpha \sqcap \beta$ denote the set of crossings of the component α with the component β. (Thus $\alpha \sqcap \beta$ does not include self-

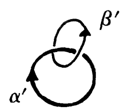
crossings of α or of β.) Then the linking number of α and β is defined by the formula:

$$lk(\alpha,\beta)=\frac{1}{2}\sum_{p\varepsilon\alpha\sqcap\beta}\varepsilon(p).$$

In other words, the linking number is one-half of the sum of crossing signs of one curve with another.

Example 1. 

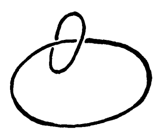
$lk(\alpha,\beta)=\frac{1}{2}(1+1)=1.$

Example 2. 

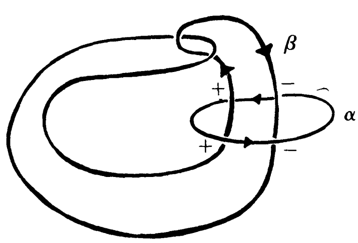
$lk(\alpha',\beta')=\frac{1}{2}(-1-1)=-1.$

Once an orientation has been assigned to a link of two components, it is immediate from the Reidemeister moves that the linking number is an invariant. For type I moves do not contribute to the linking number, while type II moves add or remove both a $+1$ *and* a -1. And type III moves do not alter the summation.

Thus the two examples above suffice to prove that the simplest link



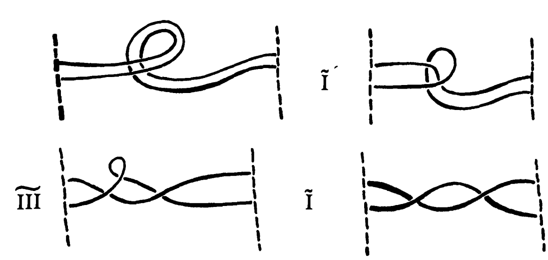
is indeed linked. For whatever orientation we assign to it, this link has a nonzero linking number.

Example 3. 

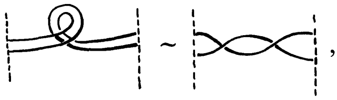
$lk(\alpha,\beta)=\frac{1}{2}(1+1-1-1)=0.$

This is the Whitehead link (named after the topologist J. H. C. Whitehead). Even though it has linking number zero, it is linked.

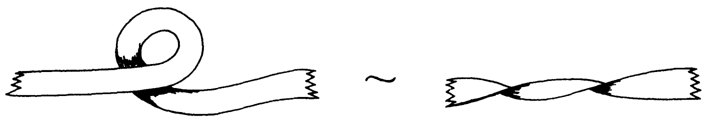
Example 4.



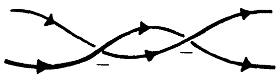
Thus



keeping the endpoints fixed. This is a familiar phenomenon that you can illustrate with a belt (the two arcs forming the edges of the belt.)

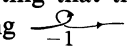


If the edges of the belt are oriented in the same direction then we can see what the linking number contribution will be from either of these forms.

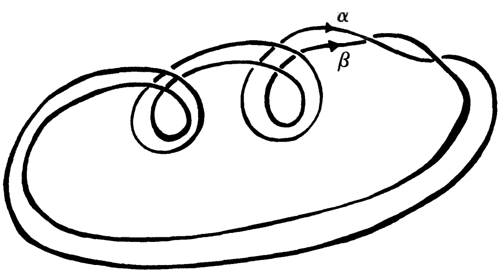


$\frac{1}{2}(-1-1)=-1$

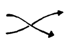
$\frac{1}{2}(-1-1)=-1$
(self-crossings are not counted)

In the curl form  it is worth noting that the linking contribution is the same as the contribution of the self-crossing .

We can use these observations to find the linking number of a more complex link such as

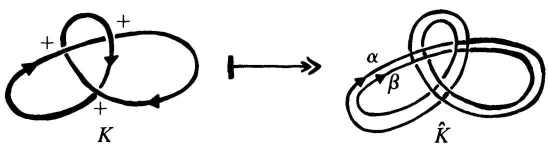


 each curl contributes $+1$ to the linking number

 occurs *twice* giving $\frac{1}{2}(1+1)=1$

$\therefore \quad lk(\alpha,\beta)=1+1+1=3.$

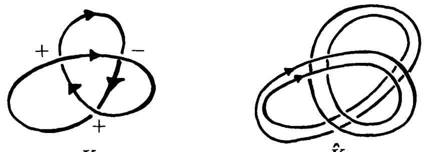
In fact, we can always find the linking number of a link that is built from a knot diagram by adding a parallel strand:



$$lk(\alpha,\beta)=w(K)=3.$$

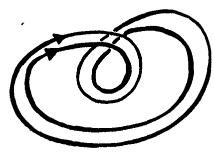
The resulting link has linking number equal to the *sum of the crossing signs of* K. We have denoted this sum by $w(K)$. It is called the *writhe of* K (or the twist number of K). The writhe $w(K)$ is not necessarily an invariant of K since it changes by ± 1 under the type I move. But the writhe *is* an invariant of the associated link $\hat{K}$ obtained by drawing parallel strands as above.

Example 5.

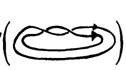


K $\qquad\qquad$ $\hat{K}$

$$lk(\hat{K}) = w(K) = +1$$

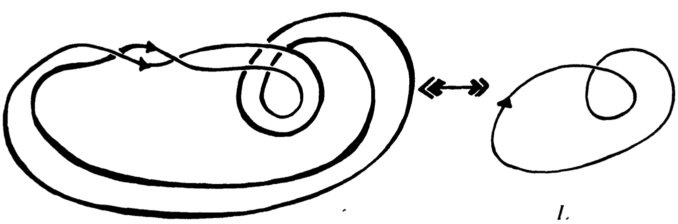
$\hat{K} \sim$  $\sim$ 

In the context of the associated parallel link $\hat{K}$, it is appropriate to call $w(K)$ *the writhe of* $\hat{K}$ and to reserve the word *twist* and a number $T(\hat{K})$ for the twisting of the strands. Thus we shall call  *one full positive twist*. And we'll write

$T($  $) = +1.$

Then for links $\hat{L}$ composed of parallel twisted strands we have the formula $lk(\hat{L}) = w(L) + T(\hat{L})$. (See [Wh].)

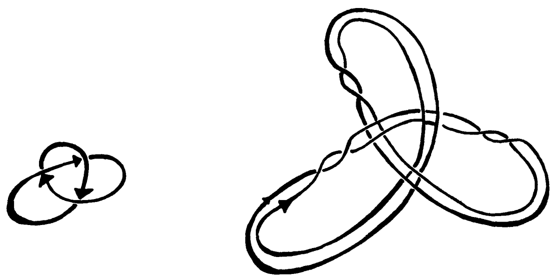
The linking number for parallel twisted strands is the sum of the writhing and twisting. Thus



$\hat{L}$ $\qquad\qquad$ L

$$\left.\begin{aligned} w(L) &= +1 \\ T(\hat{L}) &= +1 \end{aligned}\right\} \Rightarrow lk(\hat{L}) = 1 + 1 = 2.$$

Example 6.

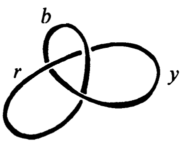


$$w(L) = +3$$
$$T(\hat{L}) = -3$$
$$lk(\hat{L}) = 3 - 3 = 0.$$

This is another example of a link with zero linking number that is nevertheless linked.

Remark. The formula [**Wh**] $lk(\hat{L}) = w(L) + T(\hat{L})$ can be regarded as a kind of "conservation law" for links of closed parallel (twisted) strands. Neither $w(L)$ nor $T(\hat{L})$ are individually topological invariants. But since the sum of writhing and twisting *is* a topological invariant, this sum must remain a constant. You can observe this conservation by playing with a rubber band. And it has been used to help understand the geometry of closed double-stranded DNA. ([**BCW**], [**F**])

There is much more to say about linking numbers—many other points of view. See *Knots and Links* by Dale Rolfsen or my books *Formal Knot Theory*; *On Knots*; *Knots and Physics* for other points of view. We now pass on to the next problem: *to show that the trefoil*  *is indeed knotted.* The most elementary proof of this fact that is known to me runs as follows: Color the arcs of the trefoil diagram red (r), blue (b) and yellow (y).



Say that a knot diagram is *tricolored* if every arc is colored r, b, or y and at any given crossing *either* all three colors appear *or* only one color appears. Of course to be tricolored there must be arcs of each color in the diagram. Then prove (exercise!) that for knots (one component) tricoloration is preserved under the Reidemeister moves.

Since the trefoil is tricolored and the unknot is not tricolored, this method articulates a topological property of the trefoil and shows that it can not be unknotted. In the next section we will show that the trefoil is *chiral*. That is, we will

prove that it is not equivalent to its mirror image. In the old days (before 1984) this was something that required a lot of mathematical background. Now we can prove it using only diagrams and a few definitions and calculations. That new invariants can be both simple and powerful makes the subject of knots very exciting.

In the next section we construct the bracket invariant and show how it gives rise to the Jones polynomial and to chirality for the trefoil. Section 3 uses the bracket to get at subtle facts about alternating knots and links. Section 4 gives more discussion of the bracket and its relation to braids and the algebra of Jones' original representation. Section 5 discusses 2-variable generalized polynomials and the historical background of Alexander and Conway polynomials. Section 6 shows how the bracket (hence, the Jones polynomial) is directly related to the Potts model in statistical physics. Sections 7 and 8 explain and generalize a relation with the Tutte polynomial in graph theory (see [T1] by M. Thistlethwaite). Section 9 discusses the knot theory of graphs embedded in three-dimensional space. Section 10 has speculations and problems.

II. The Bracket Polynomial

I begin by defining a 3-variable polynomial on unoriented link diagrams. Given an unoriented link diagram K, $[K] \in Z[A, B, d]$ will denote the corresponding polynomial in commuting variables A, B, and d. The bracket polynomial satisfies the *axioms*:

Bracket Axioms

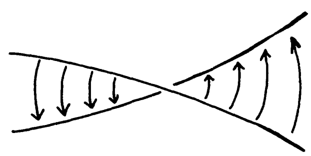
1. $[\times] = A[\asymp] + B[\,)(\,]$

   $[\times] = B[\asymp] + A[\,)(\,]$

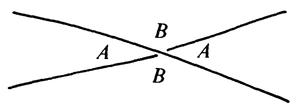
2. $[O\,K] = d[K]$

   $[O] = d.$

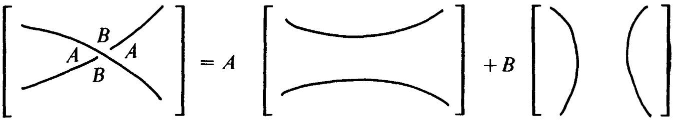
Some explanation of these rules is in order. First note that *an unoriented crossing discriminates two out of the four regions incident at its vertex.* This can be done conventionally by rotating the over-crossing line counterclockwise and choosing the two regions swept out. Thus



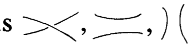
By using this convention, we can label the regions A and B, respectively:

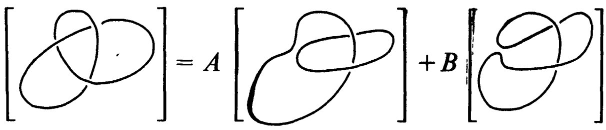


The formula in (1) then reads



and we see that A corresponds to a splice that "opens the A-channel" while B corresponds to a splice that opens the B-channel. By this convention, the second equation in (1) is correct, and a repetition of the first.

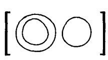
The crossings in these equations stand for larger diagrams that contain them. Thus  are assumed to be parts of *otherwise identical* diagrams. In this sense the expansion formula (1) stands for infinitely many particular formulas such as



The second equation asserts that an extra disjoint circle placed anywhere in the diagram multiplies the value of the bracket by d. In particular

$$[\text{any } N \text{ disjoint simple closed curves}] = d^N.$$

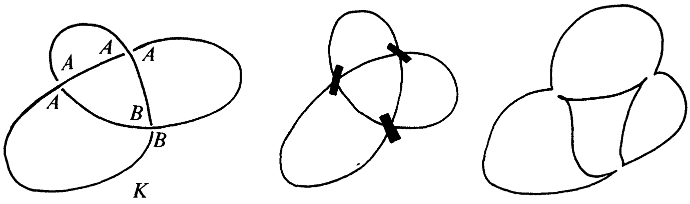
Thus

 $= d^3.$

Clearly, these axioms lead to a recursive calculation of $[K]$ by continued expansion to evaluations of collections of simple closed curves. To see that $[K]$ is well-defined, it suffices to reformulate it as a sum over *states* S of the universe U underlying K.

Let U be the universe for K. A *state S of U* is a choice of splitting for each vertex of U. I denote such a choice by a *marker at the vertex* (see FIGURE 10). FIGURE 11 shows a state of the trefoil universe and its corresponding splitting.

Given a state S, let $|S|$ denote the number of components in its splitting. Let $i_K(S)$ denote the number of A-channels opened in S and $j_K(S)$ denote the number of B-channels in S. For example,



$$i_K(S) = 2, \qquad j_K(S) = 1.$$

LEMMA 2.1. $[K] = \sum_S A^{i_K(S)} B^{j_K(S)} d^{|S|}$. *This formula for the value of the bracket follows directly from the axioms (by expanding using* (1) *and* (2)). *It gives a unique*

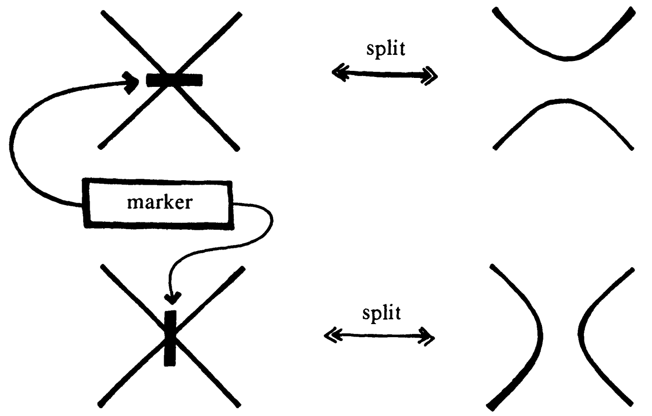


FIG. 10.

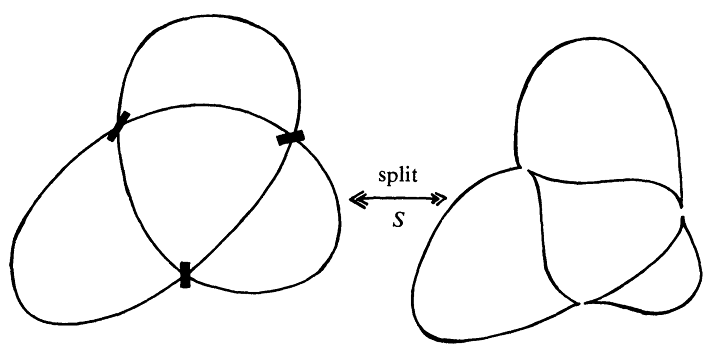


FIG. 11. A state of the trefoil universe.

value for the bracket on diagrams (no Reidemeister moves yet) and can be taken as the definition of $[K]$ *for a strictly logical development.*

We now ask: Under what restrictions on A, B, and d will $[K]$ become a topological invariant of knots and links?

This question is easy to answer via the next lemma. (See **[K5]**.)

LEMMA 2.2. $[\;\;] = AB[\supset\subset] + (ABd + A^2 + B^2)[\;\;]$.

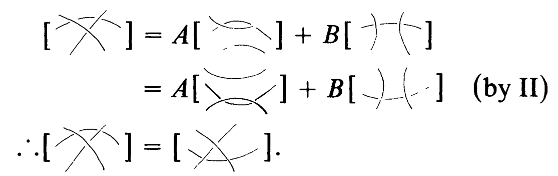
Proof.

$$[\;\;] = A[\;\;] + B[\;\;]$$
$$= A^2[\;\;] + AB[\;\;] + BA[\supset\subset] + B^2[\;\;]$$
$$= AB[\supset\subset] + (A^2 + B^2 + dAB)[\;\;].$$

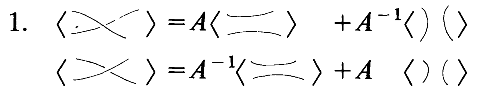
Thus with $AB = 1$ and $d = -A^2 - B^2$ we obtain invariance under the second Reidemeister move.

LEMMA 2.3. *If* [ ] = [ ], *then* [ ] *is also invariant under the type III move.*

Proof.



So for the rest of this section we'll take $B = A^{-1}$, $d = -A^2 - A^{-2}$ and also write $\langle K \rangle = d^{-1}[K]$ so that $\langle O \rangle = 1$. We then have

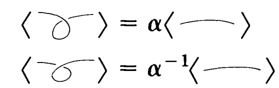


2. $\langle OK \rangle = d\langle K \rangle$

$\langle O \rangle = 1.$

This special bracket is invariant under moves II and III. It behaves as follows under the type I move:

LEMMA 2.4. *Let* $\alpha = -A^3$. *Then*



We can do two things at this point. We can understand that $\langle K \rangle$ is a special kind of invariant *and* it is possible to create an ambient isotopy invariant from $\langle K \rangle$ for K oriented. First we call the equivalence generated by moves II and III *regular isotopy*. Thus $\langle K \rangle$ is a regular isotopy invariant.

Recall from section 1 that the twist number $w(k)$ for an oriented link K is also a regular isotopy invariant. (Recall that $w(K)$ is the sum of all crossing signs.) Thus *if K is oriented* we define $f_K = \alpha^{-w(K)}\langle K \rangle$ where $\langle\ \rangle$ forgets the particular orientation. Then *f_K is an ambient isotopy invariant for oriented knots and links K.*

For mirror images we have:

LEMMA 2.5. *Let K^* denote the mirror image of K obtained by reversing all the crossings. Then* $\langle K^* \rangle(A) = \langle K \rangle(A^{-1})$ *and* $f_{K^*}(A) = f_K(A^{-1})$.

We omit the (easy) proof.

Now it turns out that f_K is a version of Vaughan Jones' original polynomial. To see this we need a definition of the Jones polynomial. Later we will have a deeper look at this. For now it suffices to say that Jones' polynomial $V_K(t)$ is determined

by the axioms: (see **[J1]**, **[J2]**, **[J3]**.)

1) $t^{-1}V \; - tV \; = \left(\sqrt{t} - \frac{1}{\sqrt{t}}\right) V$

2) $V = 1$

3) $V_K(t)$ is an invariant of ambient isotopy.

LEMMA 2.6. $f_K(t^{-1/4}) = V_K(t)$.

Proof.

$$\langle \; \rangle = A\langle \; \rangle + A^{-1}\langle \; \rangle$$

$$\langle \; \rangle = A^{-1}\langle \; \rangle + A\langle \; \rangle$$

$$\therefore A^{+1}\langle \; \rangle - A^{-1}\langle \; \rangle = (A^2 - A^{-2})\langle \; \rangle$$

$$A\alpha\langle \; \rangle\alpha^{-w(\;)} - A^{-1}\alpha^{-1}\langle \; \rangle\alpha^{-w(\;)} = (A^2 - A^{-2})\langle \; \rangle\alpha^{-w(\;)}$$

$$A\alpha f \; - A^{-1}\alpha^{-1} f \; = (A^2 - A^{-2}) f \quad (f_K = \alpha^{-w(K)}\langle K\rangle)$$

$$-A^4 f \; + A^{-4} f \; = (A^2 - A^{-2}) f$$

Let $A = t^{-1/4}$. Then

$$t^{-1} f \; - t f \; = \left(\sqrt{t} - \frac{1}{\sqrt{t}}\right) f \qquad \text{QED}$$

Having constructed the bracket, here are some sample computations:

1. $$\langle \; \rangle = A\langle \; \rangle + A^{-1}\langle \; \rangle = A(\alpha) + A^{-1}(\alpha^{-1}) = -A^4 - A^{-4}.$$

2. $$\underset{T}{\langle \; \rangle} = A\langle \; \rangle + A^{-1}\langle \; \rangle = A(-A^4 - A^{-4}) + A^{-1}(-A^{-3})^2 = -A^5 - A^{-3} + A^{-7}$$

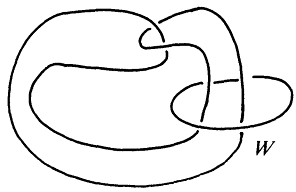
$$f_T = \alpha^{-3}\langle T\rangle = -A^{-9}\langle T\rangle = A^{-4} + A^{-12} - A^{-16}.$$

Thus $f_T(A) \neq f_T(A^{-1})$ and, hence, the trefoil knot is chiral. There is no ambient isotopy of the trefoil to its mirror image. This is the simplest known proof of the chirality of the trefoil knot. Note that all the machinery was developed from scratch, and it is all elementary.

By Lemma 2.6 we have the Jones polynomial for the trefoil as well:

$$V_T(t) = t + t^3 - t^4.$$

3. *Exercise*. Calculate $\langle W \rangle$ and show that it is a nontrivial link.



III. Alternating Knots and Links

The bracket polynomial can be used to get at some subtle properties of alternating knots and links. This comes about because we can determine a specific formula for the terms of highest and lowest degree in A for such links.

Recall that a link is said to be *alternating* if it has an alternating diagram. This is a diagram where the crossings alternate under-over-under-over-... as one travels along the link (crossing at the crossings). Now view FIGURE 12. It should be clear from this figure that in a checkerboard shading of the diagram for an alternating link, the shaded regions at each crossing are all of the same type (A or B where this is the same discrimination that we used to define the bracket). It is assumed that the underlying universe for the diagram is connected.

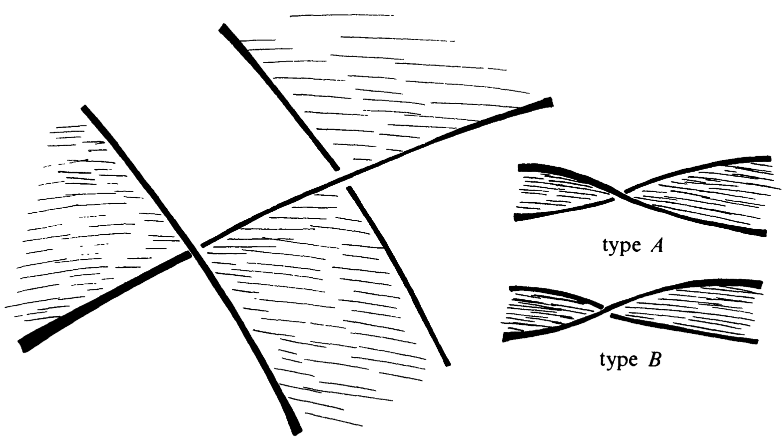
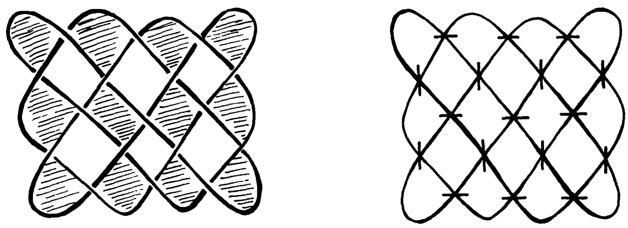


FIG. 12.

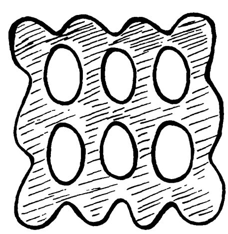
We will use this observation to guess the highest degree term in $\langle K \rangle$, and then prove that our guess is correct. Recall that the bracket is given by a summation

$$\langle K \rangle = \sum_S A^{i_K(S)} A^{-j_K(S)} d^{|S|-1}.$$

My guess is that *the highest degree is contributed by that state S where all the markers open A-channels*. Such a state will contribute a term of the form $A^V d^{|S|-1}$ where v is the number of vertices (crossings) in the diagram K. And our checker board observation shows that in the case of alternating links this *A-channel state S has W components* ($|S| = W$) where W is the number of white (unshaded) regions, and all the A-channels are colored black. View FIGURE 13.



K $\qquad$ S



S (split)

$V = 17$
$W = 7,\ B = 12$
$R = 7 + 12 = 19 = V + 2$
$|S| = 7 = W$

FIG. 13. The A-channel state.

We see that $|S| = W$ exactly because we have split all the shaded crossings, connecting the shaded part into one big shaded region, whose boundary components are boundaries of the white regions.

Thus S contributes the term

$$A^V d^{W-1} = A^V(-A^2 - A^{-2})^{W-1}.$$

Hence we assert

THEOREM 3.1. *Let K be a reduced alternating diagram. Then the highest degree term in $\langle K \rangle$ has degree given by the formula*

$$\max\deg\langle K \rangle = V + 2(W - 1),$$

where V is the number of vertices in the diagram, W is the number of unshaded regions (shading corresponding to type A crossings). This term has coefficient equal to ± 1 *in* $\langle K \rangle$. *The term of minimal degree is also monic and has degree*

$$\min\deg\langle K \rangle = -V - 2(B - 1),$$

where B is the number of shaded regions.

Comment. A diagram is *reduced* if it has no crossing that is an *isthmus*. A crossing is said to be an isthmus if any two of the four local regions at the crossing are parts of the same region in the whole diagram. See FIGURE 14.

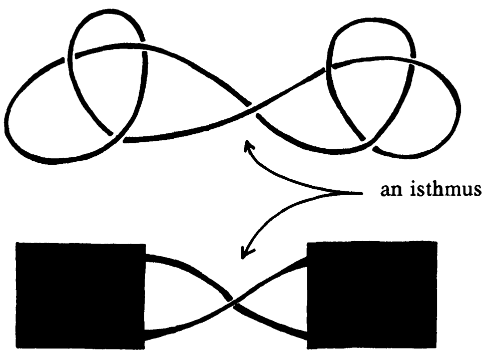


FIG. 14.

Proof of Theorem 3.1. Let S be the A-channel state. Note that any other state S' can be obtained from S by flipping some subset of S's markers. For any state S', let $\langle K|S'\rangle = A^{i_K(S')}B^{j_K(S')}d^{|S'|-1}$ denote the contribution of this state to the bracket summation. ($B = A^{-1}$, $d = -A^2 - A^{-2}$). Thus $\langle K\rangle = \Sigma_{S'}\langle K|S'\rangle$.

Now observe the following facts:

(i) If S' is obtained from S'' by flipping an A-channel marker to a B-channel marker, then $\max\deg\langle K|S'\rangle \leqslant \max\deg\langle K|S''\rangle$. The inequality is strict exactly when S' has fewer components than S''. That is, when $|S'| = |S''| - 1$.

(ii) If S' is obtained from the A-channel state by *one* flip, then $|S'| = |S| - 1$.

Assertion (i) is obvious, for if $\langle K|S''\rangle = A^x d^{|S''|-1}$, then $\langle K|S'\rangle = A^{x-2}d^{|S'|-1}$. Since S' is obtained from S'' by one flip, we know that $|S'| = |S''| \pm 1$. If $|S'| = |S''| + 1$, then

$$\langle K|S'\rangle = A^{x-2}d^{|S''|},$$

hence, $max\, deg\langle K|S'\rangle = max\, deg\langle K|S''\rangle$. If $|S'| = |S''| - 1$, then

$$\langle K|S'\rangle = A^{x-2}d^{|S''|-1-1}$$

and (using $d = -A^2 - A^{-2}$),

$$max\, deg\langle K|S'\rangle = max\, deg\langle K|S''\rangle - 4.$$

This verifies assertion (i).

Assertion (ii) is a consequence of our hypothesis of no isthmus. For suppose that $|S'| = |S| + 1$. Begin tracing along one of the components of S' at the changed marker. Note that due to our construction of the state S, this tracing (when drawn parallel to the component *in* the white regions) will encircle all or part of the original knot diagram. If $|S'| = |S| + 1$ then the two cusps $()\langle\rangle$ at the site of the

changed marker will lie on separate components of S. Thus we will end up encircling a part of the diagram, showing that this site was an isthmus. This is a contradiction. Hence, $|S'| = |S| - 1$ and $\max\deg\langle K|S'\rangle = \max\deg\langle K|S\rangle - 4$.

It follows from (i) *and* (ii) that $\max\deg\langle K|S''\rangle < \max\deg\langle K|S\rangle$ for all states S''. Thus

$$\begin{aligned}\max\deg\langle K\rangle &= \max\deg\langle K|S\rangle\\ &= V + 2(W-1).\end{aligned}$$

This completes the proof of the theorem.

We are now in a position to deduce the following (see **[K5]**, **[M2]**, **[T1]**.)

THEOREM (KAUFFMAN-MURASUGI-THISTLETHWAITE). *The number of crossings in a reduced alternating projection of a link L is a topological invariant of L.*

Proof. Let $\operatorname{span}(L)$ denote the difference between the maximal and minimal degrees of $\langle L\rangle$. Since $f_L = \alpha^{-w(L)}\langle L\rangle$ is an ambient isotopy invariant of L, we conclude that $\operatorname{span}(L)$ is also an ambient isotopy invariant. By 3.1

$$\begin{aligned}\max\deg\langle L\rangle &= V + 2(W-1)\\ \min\deg\langle L\rangle &= -V - 2(B-1),\end{aligned}$$

where V is the number of crossings in the diagram, W the number of white regions, B the number of black regions. (In a shading where all A-type crossings are shaded.) Thus

$$\begin{aligned}\operatorname{span}(L) &= V + 2(W-1) - (-V - 2(B-1))\\ &= 2V + 2(W + B - 2).\end{aligned}$$

But $W + B = R$, the total number of regions in the diagram, and $R = V + 2$. Hence, $\operatorname{span}(L) = 4V$. This completes the proof.

Discussion. This result is one of a number of classical conjectures about alternating knots and links that go back to the original compilations of knot tables by Tait and Little at the end of the last century. They also conjectured that a reduced alternating projection is minimal in the sense that it has the least number of crossings of any projection of that link. This is also true, as we shall see. Beyond this however, is the Tait *flyping conjecture*. This states that any two reduced alternating projections of the same (up to ambient isotopy) link can be obtained from another by *flyping*. A flype is a move on a tangle (with two inputs and two outputs) obtained by rotating the tangle by 180°. See FIGURE 15. Among other things, the flyping conjecture implies that the twist number, $w(K)$, of a reduced alternating projection is an ambient isotopy invariant of K. At this writing, the full conjecture remains open. Morwen Thistlethwaite *has* proved that $w(K)$ is an ambient isotopy invariant for reduced alternating diagrams. His proof uses my extension of the Brandt-Lickorish-Millett-Ho polynomial to two variables. (See section 5 for a discussion of this polynomial.)

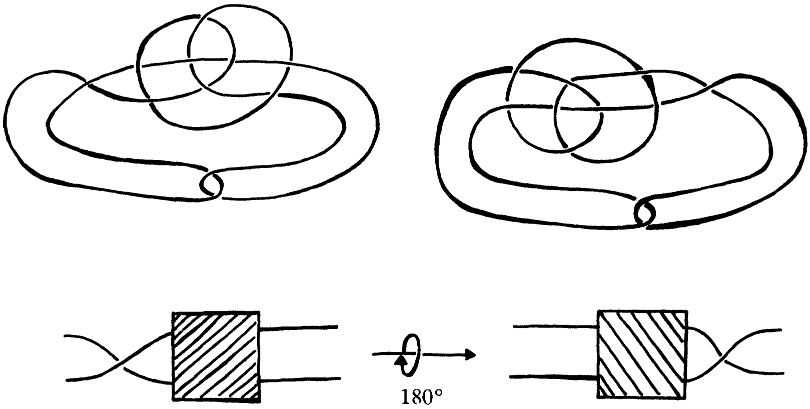


FIG. 15. Flyping.

The next lemma [K5] gives a quick proof of the general inequality span$(K) \leqslant 4V$ (first proved independently by Murasugi and Thistlethwaite).

LEMMA 3.3. *Let S be any state of a universe U. Then $|S| + |\hat{S}| \leqslant R$ where R is the number of regions in U and $\hat{S}$ is the dual state for S obtained by reversing all the markers of S.*

I omit the proof of this lemma. See FIGURE 16 for an illustration.

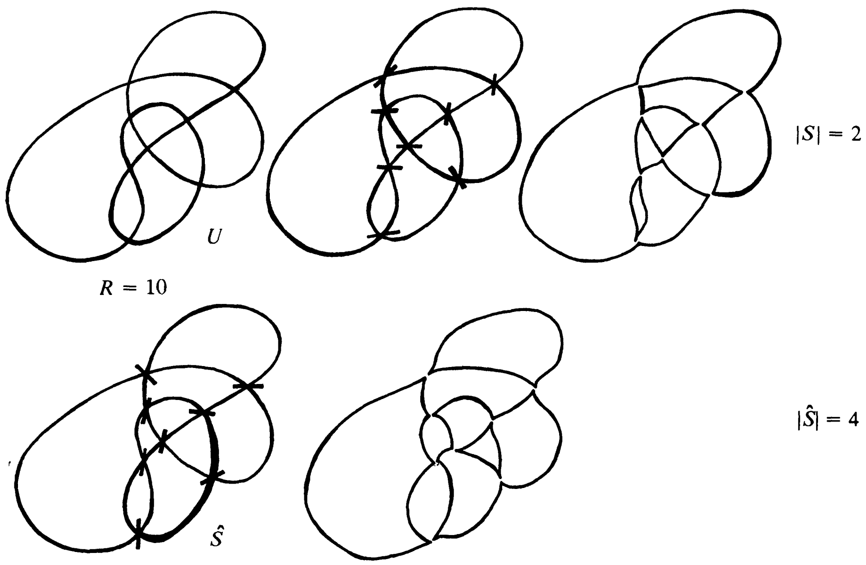


FIG. 16. $|S| + |\hat{S}| = 6 < 10$.

We then use Lemma 3.3 to prove

PROPOSITION 3.4. *For any reduced diagram* K, $\operatorname{span}(K) \leqslant 4V$, *where* V *is the number of crossings in* K.

Proof. Let S be that state for K such that every crossing is split in the A-direction. Then the same argument as in the proof of 3.1 shows that

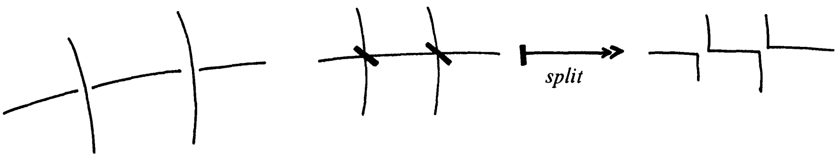
$$\max\deg\langle K\rangle \leqslant V + 2(|S| - 1)$$

and

$$\min\deg\langle K\rangle \geqslant -V - 2(|\hat{S}| - 1).$$

Therefore, $\operatorname{span}(K) \leqslant 2V + 2(|S| + |\hat{S}| - 2) \leqslant 4V$.

A refinement of Lemma 3.3 by Wu **[Wu]** shows that $\operatorname{span}(K) < 4V$ when K is any reduced nonalternating diagram. This gives us a quick proof of this inequality, also due to Murasugi and Thistlethwaite. Mr. Wu's very nice observation is that in a state S where every crossing is split in the A-direction for a nonalternating diagram there must appear splits in the pattern



corresponding to two consecutive over or under crossings. Call this a pair of parallel markers. Mr Wu notes that if S has at least one pair of parallel markers, then $|S| + |\hat{S}| \leqslant V$. Repeating the argument of 3.4, we obtain the strong inequality $\operatorname{span}(K) < 4V$ when K is reduced and nonalternating. Thus we now know that *a reduced alternating projection has a minimal number of crossings among all diagrams for the link.*

This is a remarkable application of these techniques. It is the first result of this kind in knot theory, and has a number of ramifications. For example, D. W. Sumners has used it to show that the number of knots grows at least exponentially as a function of minimal crossing number. See also **[Ki]**.

Mirror Images. We now turn to the consequences of Theorem 3.1 for chirality of alternating links. Let K be a reduced alternating projection as in 3.1. Then $\max\deg\langle K\rangle = V + 2(W - 1)$ and $\min\deg\langle K\rangle = -V - 2(B - 1)$. Thus (using $f_K = \alpha^{-w(K)}\langle K\rangle$) we have

$$\max\deg f_K = -3w(K) + V + 2(W - 1)$$
$$\min\deg f_K = -3w(K) - V - 2(B - 1).$$

If K is ambient isotopic to its mirror image K^*, then $f_{K^*}(A) = f_K(A^{-1})$ implies $f_K(A) = f_K(A^{-1})$. Hence $-\min\deg f_K = \max\deg f_K$, thus

$$3w(K) + V + 2(B - 1) = -3w(K) + V + 2(W - 1).$$

Therefore,

$$6w(K) = 2(W - B),$$

or

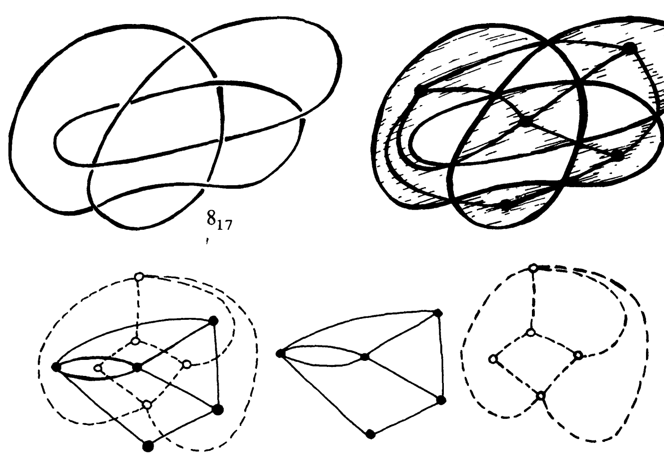
$$3w(K) = W - B.$$

Thus we have a necessary condition for an alternating link to be achiral (equivalent to its mirror image). You can check that it follows from this equation that *if the absolute value of the twist number*, $|w(K)|$, *is greater than or equal to one-third the number of crossings, then the link is chiral.* This is a step in the direction of the

TAIT CONJECTURE. *K reduced, alternating, $w(K) \neq 0$ implies K is chiral.* (Now proved by Thistlethwaite.)

Murasugi [M2] also proved the invariance of $w(K)$. His method is to note the *sum* $s(K)$ of the maximal and minimal degrees is an invariant of the ambient isotopy class of K. For K reduced alternating, we have (from the above) that $s(K) = -6w(K) + 2(W - B)$. He then uses another technique (the signature of knots and links) to show that $(W - B) - w(K)$ is an ambient isotopy invariant for the reduced alternating diagram K. Hence $w(K)$ must also be invariant. In section 5 we'll give another proof of the invariance of $w(K)$.

There exist many reduced alternating achiral knots with twist number zero. I *conjecture* that each such not only satisfies $W = B$, but that the graph associated to the white regions is isomorphic to the graph associated to the black regions. See FIGURE 17 for the example of the knot 8_{17}.



Isomorphic graph and dual graph

FIG. 17.

IV. Braids and Diagrams

Let's now consider the specialization of the bracket to the case of braids. The *n-strand braid group* B_n is generated by elements $\sigma_1, \sigma_2, \ldots, \sigma_{n-1}$ (and their inverses)

subject to the relations

$$\sigma_i \sigma_{i+1} \sigma_i = \sigma_{i+1} \sigma_i \sigma_{i+1}$$
$$\sigma_i \sigma_j = \sigma_j \sigma_i \quad \text{for} \quad |i - j| > 2.$$

The meaning of these generators and relations should become clear from FIGURE 18. A braid is a collection of unknotted strands, proceeding downward from n points (top row) to n points (bottom row). The strands wind around one another throughout the descent. Given a braid $b \in B_n$, its *closure* $\bar{b}$ is the knot or link obtained by attaching the n points in the top row to their counterparts on the bottom row. (See FIGURE 18.) By definition, the value of the bracket on a braid is its value on the closure of the braid: $\langle b \rangle = \langle \bar{b} \rangle$. Braids b_1, b_2 are multiplied by attaching the n points on the bottom row of the first to the n points on the top row of the second. (See FIGURE 18.)

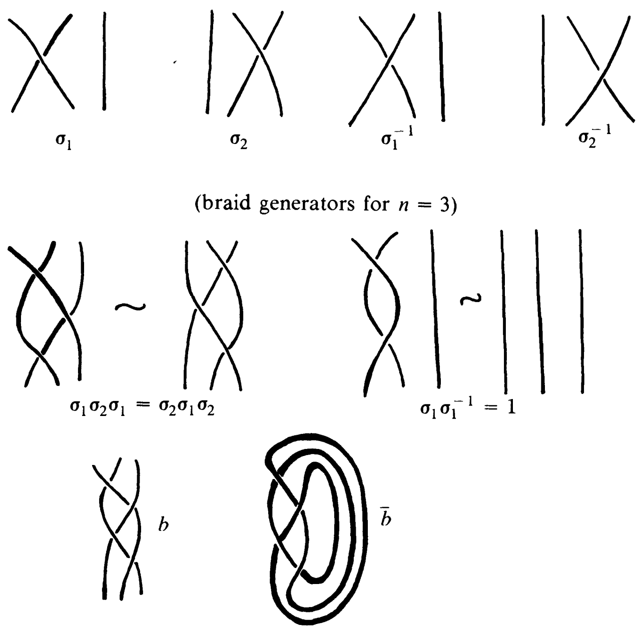


FIG. 18.

Now consider the states of a braid-universe. It should be apparent from FIGURE 19 that these can be constructed as diagrammatic products of the elementary diagrams $h_1, \ldots, h_{n-1}$ with relations

$$h_i^2 = dh_i$$
$$h_i h_{i\pm 1} h_i = h_i$$
$$h_i h_j = h_j h_i \quad \text{for} \quad |i - j| > 2.$$

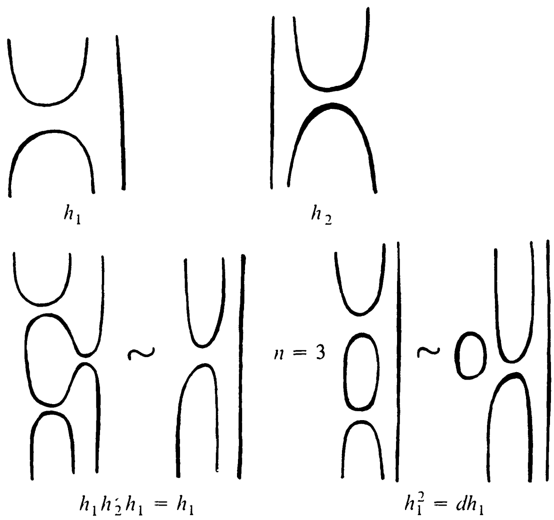
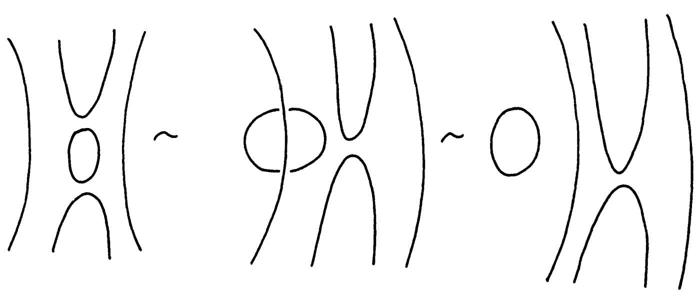


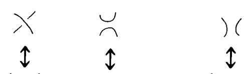
FIG. 19.

Here we take d to represent the closed loop obtained by plugging h_i into itself. As diagrams, the states can be multiplied just as we multiply braids. Since the result of such multiplications can produce extra closed loops, we need to impose a mixed topological and combinatorial equivalence relation to capture the resulting structure. Since, for computing the bracket it is irrelevant *where* a closed component is (we only count them), I define two diagrams to be *equivalent* if one can be obtained from another by regular isotopy relative to the endpoints, with free regular isotopy for closed loops. Thus,



illustrates the equivalence behind the identity $h_2^2 = dh_2$. Since the h_i's involve paired maxima and minima the first definition of this semi-group as generated by products of the h_i's is a bit unsatisfactory. We would prefer a more intrinsic definition. This can be done, but we won't go into it here (see [**K7**]). Also a mixture of braid generators and h_i's produces a more intricate structure.

Because the braid-states have a multiplicative structure, we see that the bracket expansion $\langle \times \rangle = A\langle \asymp \rangle + A^{-1}\langle) (\rangle$ can be construed *for braids* as a mapping $\rho: B_n \to D_n$ where D_n is the free additive algebra over $Z[A, A^{-1}]$ with multiplicative generators h_i and relations (*) above. That is, we define



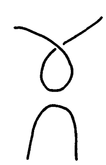
$$\rho(\sigma_i) = Ah_i + A^{-1}1$$

$$\rho(\sigma_i^{-1}) = A^{-1}h_i + A$$

and take $d = -A^2 - A^{-2}$ in D_n. Then the formalism we have used to prove that $\langle K \rangle$ is an invariant of ambient isotopy also proves (via the braiding relations) that ρ is a representation of the n-strand braid group to the algebra D_n. Furthermore there is a function tr: $D_n \to Z[A, A^{-1}]$ that we may interpret as the linear extension of tr(h) where h is a product of h_i's. And tr$(h) = d^{|h|-1}$ where $|h|$ is the number of disjoint circles in the state corresponding to h. Then $\text{tr} \circ \rho(b) = \langle b \rangle$ and *this gives a diagrammatic interpretation to the original construction of the Jones polynomial via representations.*

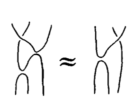
This approach has been generalized (see **[K6]**, **[K7]**, **[L]**) but the algebra of the h_i's remains the most transparent structure in this context. And while it may seem transparent, it is in fact rather opaque: We do not yet know whether there is a non-trivial knot with trivial Jones polynomial.

The Mixed Algebra. From our context, it is very natural to consider a mixture of products of braid generators σ_i and the state-elements h_i. At this writing, an abstraction of this algebra has been used by Birman and Wenzel **[BW]**, and one very beautiful systematization of it by Yetter **[Y]**. For our purposes we shall write such diagrams up to regular isotopy. Thus we do not have relations $\sigma_i h_i = h_i$ since (for example) $\sigma_1 h_1$ has diagram



and it requires a type I move to obtain the cancellation.

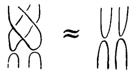
And since diagram multiplication does yield an extra loop d when squaring h_i, we hope to retain the relation $h_i^2 = dh_i$. Let $\mathcal{M}_n$ denote this (multiplicative) extension of braids via the h_i's. Obviously, we want a better formal definition of $\mathcal{M}_n$, but first consider some examples:

1°.  $\sigma_1 h_2 h_1 = \sigma_2^{-1} h_1.$

This is a fundamental type of mixed relation. Note how the pairing of

maxima/minima to produce the h_1 on the left-hand side comes from different arcs than on the right-hand side!

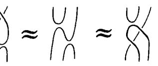
2°.



$$\sigma_2\sigma_3\sigma_1\sigma_2 h_1 h_3 = h_1 h_3.$$

We are allowing regular isotopy of the strands relative to the endpoints and to the (vertical) sides of the box in which the tangle sits.

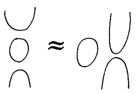
3°.



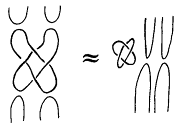
$\sigma_2\sigma_1 h_2 = h_1 h_2 = h_1\sigma_2\sigma_1.$

It may begin to look like there is a myriad of possible relations in $\mathscr{M}_n$. This is true, but *the kind of relation illustrated in example* 1° *plus the usual braiding and* h_i *relations is sufficient to generate the others*. (For details see **[Y]** and compare with **[K6]** and **[K7]**.) For example:

Thus one can actually give generators and relations for $\mathscr{M}_n$, just as for the classical braid group. But in order to do so a decision must be made about handling appearances of closed loops. We take

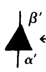


bringing closed loop forms outside the rectangle, then it is natural to move whole knots and links in and out as in:

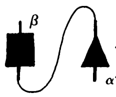


Thus in this formulation, we can write $h_1 h_3 \sigma_2 \sigma_1^{-1} \sigma_3^{-1} \sigma_2 h_1 h_3 = \Lambda h_1 h_3$ where Λ is the given link. By allowing multiplication by disjoint union with knots and links, we go beyond a simple set of generators and relations. However, the formalism is useful in some contents. For example, using $[K] = d\langle K\rangle$ we have $[K \sqcup K'] = [K][K']$ so that the square bracket preserves this outer multiplicative structure.

Finally, the outer form of multiplication then fits in with a generalized tensor formalism (see **[P]**) with two types of multiplication corresponding to ordinary tensor product and to different forms in index contraction such as matrix multiplication. Thus if

 $\leftrightarrow A^{\beta}_{\alpha}$ and  $\leftrightarrow B^{\beta'}_{\alpha'}$

then


$\leftrightarrow A^{\beta}_{\alpha} B^{\alpha}_{\alpha'}$

(meaning the sum over all occurrences of α by the summation convention). Connection by a single connecting line corresponds to ordinary matrix multiplication. Connection using multiple connecting lines can correspond to multiplication in a tensor product.

We can regard $\cap$ and $\cup$ as diagrams for matrices $\mathscr{M}_{\alpha\beta}$ and $\mathscr{M}^{\alpha'\beta'}$ respectively. Then $\substack{\cup\\ \cap}$ corresponds to the tensor product $h = \mathscr{M}_{\alpha\beta}\mathscr{M}^{\alpha'\beta'}$, and $h^2 = \mathscr{M}_{\alpha\beta}\mathscr{M}^{\alpha'\beta'}\mathscr{M}^{\alpha\beta}\mathscr{M}_{\alpha''\beta''} = (\mathscr{M}_{\alpha\beta}\mathscr{M}^{\alpha\beta})h = \Delta h$. This formalism corresponds directly to the diagram for h^2 with $\bigcirc$ corresponding to the scalar $\Delta = \mathscr{M}_{\alpha\beta}\mathscr{M}^{\alpha\beta}$.

In this way our diagram algebra can be interpreted as the underlying structure for specific matrix representations of the multiplicative structure of the h_i's. See [**K7**] for a complete exposition of this.

In this last comment we have informally presented two points of view about the extended braid-like multiplicative structures that appear so naturally from the braid-states. By restricting to internal multiplication (matching upper and lower strands) one obtains significant generalizations of the Artin braid group. Adding outer multiplication by closed forms (knots and links) creates close correspondences with representations and abstract tensor products.

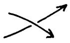
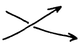
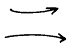
V. Generalized Polynomials

There are, at present, two 2-variable generalized polynomial invariants for knots and links, each a generalization of the Jones polynomial. These are the Homfly polynomial and the Kauffman polynomial. I will denote the Homfly polynomial by $P_K(\alpha, z)$ and the Kauffman polynomial by $F_K(\alpha, z)$. In this section we will touch on the formalisms of these polynomials. And I shall begin by recalling the Conway polynomial and telling a bit of the tale leading from Alexander to Conway to generalized polynomials.

In the beginning [**A**] was Alexander and his invention/discovery of the Alexander polynomial $\Delta_K(t)$. Alexander probably discovered this polynomial by thinking about covering spaces, but his paper was strictly combinatorial, using linear algebra, determinants, and the Reidemeister moves. He showed that if two oriented knots or links K, K' are ambient isotopic, then $\Delta_K(t) \doteq \Delta_{K'}(t)$ where $\doteq$ means equal up to a multiple of $\pm t^n$ for some integer n. The polynomial was seen to be quite good at distinguishing knots and links, although it did not distinguish a knot or link from its mirror image.

The Alexander polynomial has been an extraordinary and useful tool in knot theory since its inception. Attempts to model and reformulate it led to much new work and different points of view. One of the most notable of these approaches is R. H. Fox's [**CF**], discovery of the *free differential calculus*, a technique for

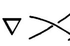
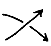
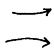
extricating the Alexander polynomial from any presentation of the fundamental group of its complement. Then in 1970 John Horton Conway published a remarkable paper [**Con**] in which he showed that the Alexander polynomial could be sharpened to an invariant with a simple recursive definition. The Conway polynomial, $\nabla_K(z)$, is determined by the conditions:

1. ∇  $-\nabla$  $= z\nabla$ 
2. ∇  $= 1$
3. $\nabla_K(z) = \nabla_{K'}(z)$ whenever K and K' are ambient isotopic.

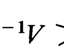
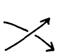
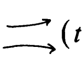
Conway explained that his polynomial was related to the Alexander polynomial by the formula $\Delta_K(t) \doteq \nabla_K(\sqrt{t} - 1/\sqrt{t})$.

Eight years later, Conway became enthusiastic once again about this polynomial and he lectured about it in a number of places. This time people heard him and their interest led to some papers about the polynomial (see e.g., [**K1**], [**G**], [**Co**]). The focus was primarily on how to use this recursive scheme, and on understanding the relation to the Alexander polynomial. Some use was made of the extra information in the Conway polynomial. (It can distinguish many links with even number of components from their mirror images.) This author wrote a monograph [**K2**] on combinatorial and diagrammatic work related to $\nabla_K(z)$. In particular, I found a states model for $\nabla_K(z)$ with a state-summation that is a bit more intricate than our model for the bracket. This model allowed a new proof of the theorems of Murasagi and Crowell ([**M1**], [**Cr**]) on the genus of alternating knots, and a generalization of these results to a category I called alternative knots.

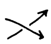
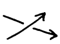
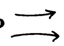
But curiously, no one tried to generalize Conway's recursive scheme itself. No one asked what would happen if the first formula were modified to, say,

∇  $+\nabla$  $= z\nabla$  ?

And then in 1984 Vaughan Jones lectured on his new invariant, derived from a representation of the Artin braid group into a von Neumann algebra [**J2**]. And Vaughan proved (among other things) that his (Laurent) polynomial satisfied an identity

$t^{-1}V$  $(t) - tV$  $(t) = \left(\sqrt{t} - \frac{1}{\sqrt{t}}\right)V$  (t).

With this formula standing in juxtaposition to the Conway formula, a number of people leapt at once to the generalization

$\alpha^{-1}P$  $-\alpha P$  $= zP$ ,

giving a two-variable polynomial $P_K(\alpha, z)$ specializing both to the Conway ($\alpha = 1$) and the Jones ($\alpha = t$, $z = \sqrt{t} - 1/\sqrt{t}$) polynomials. This is the Homfly polynomial [**HOMFLY**].

Some time passed, and then Brandt, Lickorish, Millett, and (independently) Ho found ([**BLM**], [**H**]) yet another new invariant polynomial $Q_K(z)$ satisfying

$Q + Q = z(Q + Q)$ for unoriented links. This is a one-variable polynomial, distinct from the Homfly polynomial. It does not distinguish mirror images.

I then had the good fortune to recognize how to put another variable into the context of the Q-polynomial ([**K4**], [**K6**]). The idea is to work in the regular isotopy category (as we explained for the bracket) and let a polynomial L be defined via:

1) $L + L = z(L + L)$
2) $L = \alpha L$
   $L = \alpha^{-1} L$
3) $L = 1$
4) $L_K = L_{K'}$ whenever K and K' are regularly isotopic.

Then L is normalized to form an invariant of ambient isotopy for oriented knots and links via the equation

$$F_K = \alpha^{-w(K)} L_K$$

where $w(K)$ is the twist number of the diagram K. The polynomial F_K turns out to be quite good at distinguishing knots and links from their mirror images. It appears to be a proper companion to the Homfly polynomial $P_K(\alpha, z)$. (The reader should note that our names of polynomials by letter and variable choice may differ from those given elsewhere in the literature. The translations are always straightforward.) Ocneanu and Jones discovered how to put a trace on the Hecke algebra generated by elements c_i satisfying braiding relations and the Conway-type relation

$$c_i - c_i^{-1} = z$$

to produce the Homfly polynomial in a fashion analogous to the representation for $V_K(t)$. Hugh Morton worked extensively with this algebra, producing very good programs for computing the Homfly polynomial for braids. Morton continues doing deep theoretical work related to these polynomials. (See [**M**].)

Birman and Wenzel [**BW**] have given a similar treatment for the Kauffman polynomial by using an algebra with relations $c_i + c_i^{-1} = z(1 + E_i)$, where c_i corresponds to a braid generator and E_i shows the formal properties of our h_i's. More will come of this. David Yetter [**Y**] has given a good general context for diagram-related algebras. See also [**K5**], [**K6**], and [**K7**].

Both of the two-variable generalized polynomials have the Jones polynomial as a special case. For F_K this is most easily seen via adding:

$$\langle\ \rangle + \langle\ \rangle = (A + A^{-1})(\langle\ \rangle + \langle\ \rangle)$$

$$\langle\ \rangle = \alpha\langle\ \rangle$$

$$\alpha = -A^3$$

$$\langle\ \rangle = \alpha^{-1}\langle\ \rangle$$

Thus $\langle K\rangle = L_K(-A^3, A + A^{-1})$ and $f_K = F_K(-A^3, A + A^{-1})$. Since $V_K(t) =$

$f_K(t^{-1/4})$ we conclude that $V_K(t) = F_K(-t^{-3/4}, t^{-1/4} + t^{1/4})$. This was observed by Lickorish [L] by a different route.

The two 2-variable polynomials can be established via direct inductive definition. It is an open question whether there exist models for the workings of these polynomials that connect them directly with geometry beyond the geometry of diagrams. I believe that such connections will come about, and that they will be of great importance for topology as a whole.

Remark. Thistlethwaite's simple proof [**T2**] of the invariance of the writhe for reduced alternating diagrams uses the Kauffman polynomial: He observes that for L_K with K reduced alternating, the highest term in z has coefficient $k(a + a^{-1})$, $k > 0$ and power z^{n-1} where n is the number of crossings in K. Thus $L_K = k(a + a^{-1})z^{n-1}$ + (other lower degrees in z) for K reduced alternating. Since $F_K = \alpha^{-w(K)}L_K$ is an ambient isotopy invariant, it follows at once that $w(K)$ is also an ambient isotopy invariant. The proof of Morwen's observation is a direct structural induction:

(i) If is reduced alternating, then and $)\,($ are both alternating, and at least one is reduced.

(ii) Use (i) and the recursion

$$L + L = z(L + L_{)\,(}).$$

The separation of the α and z variables is the crucial ingredient in the proof.

VI. Graphs and Statistical Physics

Recall that in section III, I defined a generalized bracket polynomial for diagrams so that $[K] \in Z[A, B, d]$ and

1. $[\;] = A[\;] + B[\;)\,(\;]$
2. $[0\ K] = d[K]$
   $[0] = d.$

We then created an invariant of regular isotopy via $B = A^{-1}$, $d = -A^2 - A^{-2}$ and $\langle K \rangle = d^{-1}[K]$ (so that $\langle 0 \rangle = 1$). The square bracket can be specialized in other ways. In particular, it is (for the right choice of A and B) the dichromatic (Whitney-Tutte) polynomial for a planar graph. This in turn can be seen to be a way of expressing the partition function for the Potts model (a generalization of the Ising model) in statistical physics [**B**].

To understand this connection it is important to realize the following:

THEOREM 6.1. *Universes are in one-to-one correspondence with planar graphs.*

Proof. To each universe, shade it so that the unbounded region is unshaded (i.e., 2-color the regions). Associate a graph to U, $\Gamma(U)$, so that the vertices of $\Gamma(U)$ correspond to the shaded regions of U and the edges correspond to crossings shared by shaded regions. Given a graph G, associate to it a universe $V(G)$ by placing a crossing of the form on each edge of G and connecting these

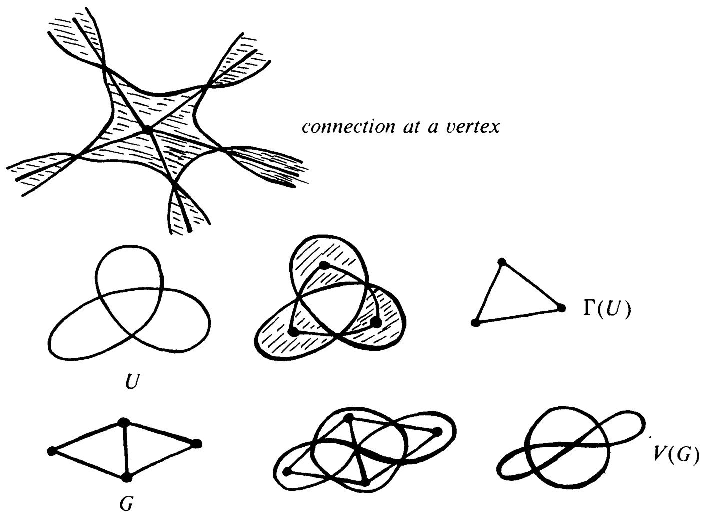


FIG. 20.

crossings at each vertex as shown in FIGURE 20. It is easy to verify that $\Gamma(V(G)) = G$ and $V(\Gamma(U)) = U$. This completes the proof.

The *dichromatic polynomial* $Z_G(q, v) \in Z[q, v]$ is defined for graphs G by the recursive formulas:

1) $Z \quad = Z \quad + vZ$
2) $Z_{\bullet G} = qZ_G$
   $Z_{\bullet} = q$

The first formula asserts that the value of Z on a graph G is equal to the sum of the value of Z on a graph G' obtained by deleting one edge from G plus v multiplied by the Z for G'', the graph obtained by collapsing this edge to a point. The second formula asserts that the addition of an extra vertex to a graph G multiplies the dichromatic polynomial by q, and that the value of Z for an isolated vertex is q.

Examples.

$$Z \quad = Z_{\bullet\ \bullet} + vZ_{\bullet} = q^2 + vq$$

$$Z \quad = Z_{\bullet} + vZ_{\bullet} = q + vq$$

$$Z \quad = Z \quad + vZ$$

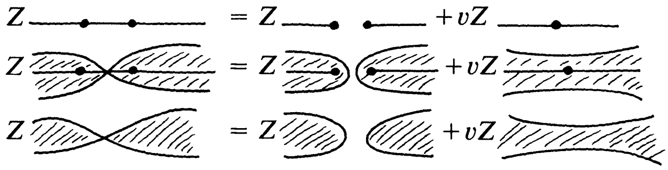
$$= Z \quad + vZ \quad + v(Z \quad + vZ \quad)$$

$$= q(q^2 + vq) + v(q^2 + vq) + v(q^2 + vq) + v^2(q + vq)$$

$$= (q + 2v)(q^2 + vq) + v^2(q + vq).$$

For $v = -1$, the dichromatic polynomial specializes to the *chromatic polynomial*. That is, $Z_G(q, -1) = K_G(q)$ is the number of ways to vertex-color the graph G with q colors so that no two adjacent vertices receive the same color. This is easily seen from the recursion formula since K counts all possibilities for these two vertices, while K corresponds to those cases where the two vertices receive the same color.

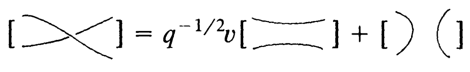
Now consider how the recursion formula (1) diagrams:



We see that deletion and contraction in the graphs become the two ways of splicing the crossing in the knot diagrams (universe). And the expansion for Z is formally a bracket expansion.

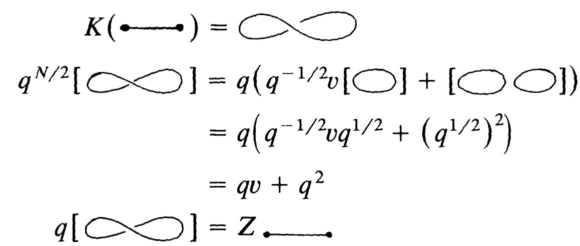
Some further translation is then required to actually re-write Z_G as a bracket. First let $K(G)$ be the *alternating* link diagram associated with $V(G)$ so that all shaded crossings are of type A. Then

THEOREM 6.2. *$Z_G(q, v) = q^{N/2}[K(G)]$ where N is the number of vertices of G, and the bracket is expanded with $A = q^{-1/2}v$, $B = 1$, $d = q^{1/2}$ so that*



and $[0] = q^{1/2}$.

Example.



Thus we see that the square bracket is fundamentally related both to knot theory and to graph theory. This connection raises many questions. We would like to know whether qualitative information can be transferred between these two subjects. The square bracket gives a picture of a parameter space A, B, d and subvarieties along which $[K]$ is topological or dichromatic. More work is needed here.

Just to complete this picture I shall explain how the partition function for the Potts model in statistical physics is a dichromatic polynomial (see [**B**], [**K7**]). The

partition function has the form

$$Z_G = \sum_{\sigma} e^{-E(\sigma)},$$

where σ runs over all "states" of the lattice G (we will let G be a planar graph) and $E(\sigma)$ is the energy of the given state. In the Potts model the energy has the form

$$E(\sigma) = \frac{1}{kT} \sum_{\langle i, j \rangle} \delta(\sigma_i, \sigma_j),$$

where $\langle i, j \rangle$ denotes an edge of G with vertices i, j and σ_i and σ_j are the state's assignments to these vertices. We assume that each vertex can be freely assigned one of q values, and that a state σ is such an assignment. In this formula δ is the Kronecker delta

$$\delta(a, b) = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{otherwise} \end{cases}$$

and T is the *temperature* of the system, while k is a constant (Boltzman's constant).

The partition function has many uses in this subject. For example, the probability of being in a given energy state E is

$$p(E) = e^{-E}/Z_G.$$

PROPOSITION 6.3. *For*

$$E(\sigma) = \frac{1}{kT} \sum_{\langle i, j \rangle} \delta(\sigma_i, \sigma_j)$$

and q local states, let

$$v = e^{-(1/kT)} - 1.$$

Then the partition function is the dichromatic polynomial in q and v:

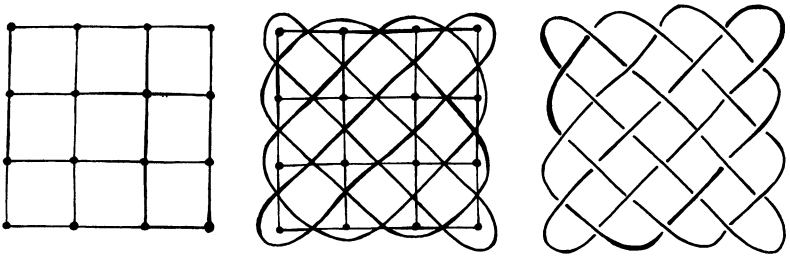
$$\sum_{\sigma} e^{-E(\sigma)} = Z_G(q, v).$$

Proof.

$$\begin{aligned} \sum_{\sigma} e^{-E(\sigma)} &= \sum_{\sigma} e^{-1/kT \sum_{\langle i, j \rangle} \delta(\sigma_i, \sigma_j)} \\ &= \sum_{\sigma} \prod_{\langle i, j \rangle} \left(e^{-(1/kT)\delta(\sigma_i, \sigma_j)}\right) \\ \sum_{\sigma} e^{-E(\sigma)} &= \sum_{\sigma} \prod_{\langle i, j \rangle} \left(1 + v\delta(\sigma_i, \sigma_j)\right). \end{aligned}$$

It is easy to see that the right hand side of this equation satisfies the recursion relation for the dichromatic polynomial in q and v.

$(Z_{\bullet G} = qZ_G)$



G K(G)

FIG. 21. Translating the square lattice.

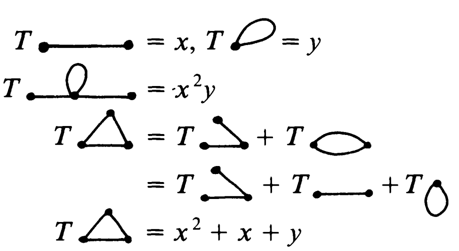
Combining 6.2 and 6.3 we see that the partition function is a bracket expansion. This gives a theoretical explanation for the appearance of the algebra of the h_i's (see section 4) in the structure of the Potts model for the square lattice. (See [**B**].) This lattice corresponds to the plat closure (see FIGURE 21) of a particular braid. Any bracket evaluation for a braid is expressed in terms of this operator algebra. It remains to be seen whether the bracket formulation for the Potts model will shed light on its physics. The relationship between the Potts model and the Jones polynomial via its operator algebra was first observed by Vaughan Jones. Our formulation shows the direct connection through translating graphs and link diagrams.

VII. The Bracket and the Tutte Polynomial

In the last section we showed how the bracket could be used to give the dichromatic polynomial for planar graphs. Here we shall reformulate the state expansion of the general bracket function, showing that it can be calculated solely from states with a single component. Our reformulation generalizes work of Morwen Thistlethwaite. He showed [**T1**] how to do this for the bracket and the Jones polynomial. The ideas go back to a generalization of the dichromatic polynomial known as the Tutte polynomial. (See [**Tu**]). The Tutte polynomial is defined recursively as follows.

To each graph G is associated a polynomial $T_G(x, y) \in Z[x, y]$. If G is composed solely of isthmus and loops then $T_G = x^i y^l$ where i is the number of isthmuses and l is the number of loops. The polynomial satisfies the recursion $T_G = T_{G'} + T_{G''}$ where G' and G'' are the graphs obtained by deleting and contracting (respectively) an edge that is neither a loop nor an isthmus.

Examples.



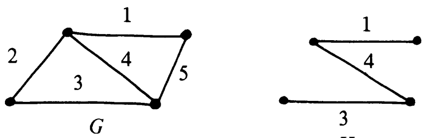
The dichromatic polynomial $Z_G(q, v)$ (see section 6) is related to the Tutte polynomial $T_G(x, y)$ by the formula $Z_G(q, v) = qv^{N-1}T_G(1 + qv^{-1}, 1 + v)$ where N is the number of vertices of G. For example, $qv(1 + q/v) = qv + q^2 = Z$ and $q(1 + v) = q + qv = Z$  show that the formula is correct for a single loop and isthmus. This formula shows that the dichromatic polynomial and the Tutte polynomial determine one another. Thus

$$T_G(x, y) = \frac{1}{(x-1)(y-1)^N} Z_G((x-1)(y-1), (y-1)).$$

Tutte proved a remarkable theorem showing that his polynomial could be computed from weightings assigned to the maximal trees of the graph G. This weighting is dependent upon an ordering of the edges of G, but independent of the particular ordering.

Definition 7.1. Let G be a graph whose edges have been labelled $1, 2, 3, \ldots, n$. Let $H \subset G$ be a maximal tree in G. Let $i \in \{1, 2, \ldots, n\}$ denote an edge of H. Let H_i denote $H -$ (the ith edge). Since H is a maximal tree, H_i has two components. One says that i is *internally active* if $i < j$ for every edge j in G-H and endpoints in both components of H_i. Let $i \in G$-H be an external edge. One says that i is *externally active* if $i < j$ for all edges j on the cycle in H extending from one end of i to the other.

Example.



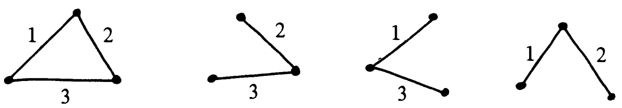
G $\qquad$ H

Here the edge labelled 1 is internally active. Edge number 2 is externally active.

THEOREM (TUTTE). *Let $\mathscr{H}$ denote the collection of maximal trees in a graph G. Let $i(H)$ denote the number of internally active edges in G, and $e(H)$ the number of externally active edges in G (with respect to a given tree H). Then the Tutte polynomial is given by the formula*

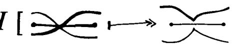
$$T_G(x, y) = \sum_{H \subset \mathscr{H}} x^{i(H)} y^{e(H)}.$$

Example.

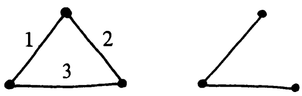


G $\qquad$ H_1 $\qquad$ H_2 $\qquad$ H_3

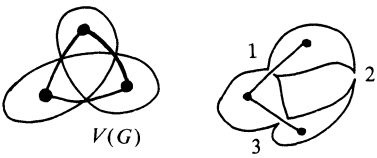
$$T_G = y + x + x^2.$$

It may seem from the definitions of internal and external activity that they are somewhat different. Actually there is a symmetry of definition for planar graphs. It is through this symmetry that I like to see the relationship of the Tutte weightings with universes and with knot theory. To see this symmetry consider the universe $V(G)$ (sometimes called the medial graph) associated with a planar graph G. Each maximal tree $H \subset G$ determines a state $S = S(H)$ of $V(G)$ with one component. This state is obtained by splitting $V(G)$ along the edges of H [], and splitting all other crossings in the opposite fashion.

Example.



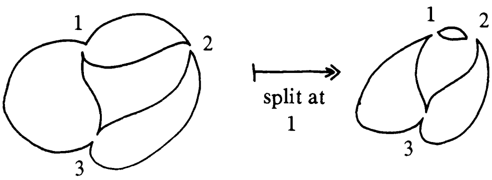
G H



$S(H)$

The edge labelling of G becomes a vertex labelling of $V(G)$. Call the vertices (crossings) of $S(H)$ *internal* or *external* according as they are split with cusps pointing to the *inside* or to the *outside* of the Jordan trail $S(H)$. Thus in the example above, 1 and 3 are internal and 2 is external. Given a Jordan trail S (a universe with $|S| = 1$) with vertices (*sites* $\times$) labelled $1, 2, \ldots, n$, call a site i *active* if $i < j$ for all sites j with cusps in the two components resulting from splitting S at i. (Compare [**K2**].)

Example.

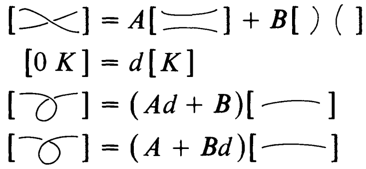


Since $1 < 2$, 1 is active.

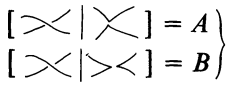
A site on a trail is *internally active* if it is internal and active. A site is externally active if it is external and active. By replacing the trees in G by Jordan trails on $V(G)$, we obtain a symmetrical definition of $T_G(x, y)$.

Incidentally, it is easy to see from this reformulation that $T_G(x, y) = T_{\overline{G}}(y, x)$ where $\overline{G}$ is the planar dual graph to the planar graph G. Each Jordan trail gives a pair of maximal trees, one for G and one for $\overline{G}$.

By now we are very close to the knot theory, and I can explain how to calculate the square bracket, $[K]$ for link diagrams by using Tutte weightings. Recall that $[K]$ has variables A, B, d and that

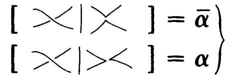


Let $\alpha = Ad + B$ and $\bar{\alpha} = A + Bd$. Now let K be a given diagram, and S be a state for K with one component ($|S| = 1$). Let $\mathcal{S}$ denote the collection of all states S with one component. Let the crossings of K be labelled $1, 2, \ldots, n$. Each crossing of K will determine a local contribution at the corresponding site of S. If the site is *inactive* we retain the usual bracket contribution:



inactive site.

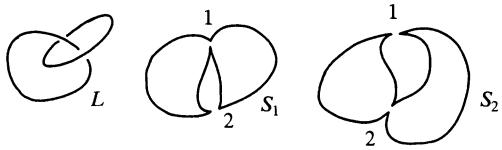
If the site is active then we take



active site.

Then $[K|S]$ is the product of all of these local contributions, and we assert that $[K]$ is given by the formula $[K] = \Sigma_{S \in \mathcal{S}}[K|S]$.

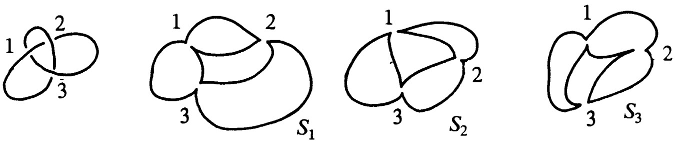
Example 1.



$[L] = \Sigma_{\mathcal{S}}[L|S] = \bar{\alpha}B + \alpha A$ (1 active in S_1, 1 also active in S_2). Doing this the long way we have,

$$= A(Ad + B) + B(A + Bd)$$
$$= A\alpha + B\bar{\alpha}.$$

Example 2.



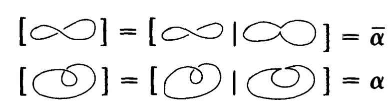
$$[K] = \sum_{\mathcal{S}}[K|S] = \bar{\alpha}BA + \alpha A^2 + \bar{\alpha}^2 B.$$

Note that for $\langle K \rangle$, $\alpha = -A^3$, $\bar{\alpha} = -A^3$, $B = A^{-1}$ so that

$$\langle K \rangle = -A^{-3} - A^5 + A^{-7},$$

our familiar value for the bracket of the trefoil. (See section 2.)

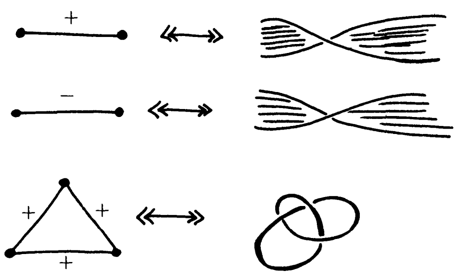
Example 3.



These correspond graphically to basic contributions for isthmus and loop. Note however, that if we switch the crossing then the contributions flip (since we are cataloging the type of curl in the knot diagram).

Warning. It should be clear to both author and the reader by now that the square bracket function we are now using is normalized to 1 on the circle: [0] = 1. Let this cause no difficulty with regard to our earlier convention.

Example 4. This is really a reformulation. Note that diagrams of knots and links are in 1-1 correspondence with *signed planar graphs* where the signs are placed on the edges of the graph so that + corresponds to an A-channel, $-$ corresponds to a B-channel.

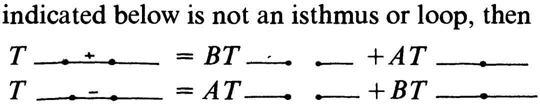


Our Tutte-reformulation of the generalized bracket then gives a *generalized Tutte polynomial for signed graphs* satisfying

1) If G has i_+ positive isthmuses, i_- negative isthmuses, l_+ positive loops, l_- negative loops, then

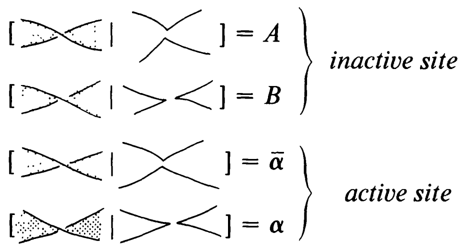
$$T_G = X^{i_+ + l_-} Y^{i_- + l_+}.$$

2) If the edge indicated below is not an isthmus or loop, then



Thus $T_G(A, B, x, y)$ is a graph theoretic version of the square bracket. ($x = \alpha$, $y = \bar{\alpha}$ recovers previous notation). It has a Tutte expansion in terms of spanning trees, and should be explored for its own sake. (*N.B.* $Ax + B^2 = A^2 + By$).

Example 5. If K is an alternating diagram then all crossings have the same internal type. Thus the contributions take the form



From this it is easy to see some specifics about the topological bracket where $B = A^{-1}$, $\alpha = -A^3$, $\bar{\alpha} = -A^3$. Note that $\alpha = -A^4(A^{-1})$, $\bar{\alpha} = -A^{-4}(A)$. Thus for K alternating, we have

$$\langle K \rangle = A^{I-E} T_{G(K)}(-A^{-4}, -A^4),$$

where T is the (standard) Tutte polynomial and I denotes the number of internal sites on a trail, E the number of external sites on a trail. We note $I - E$ is a constant independent of the given choice of trail.

Now recall that the ambient isotopy invariant f_K is given by the formula

$$f_K = \alpha^{-w(K)}\langle K \rangle = (-A^{-3})^{w(K)}\langle K \rangle$$

and that the Jones polynomial $V_K(t)$ is given by the formula

$$V_K(t) = f_K(t^{-1/4}).$$

Thus for K alternating we have:

THEOREM (THISTLETHWAITE). *Let K be an alternating projection, $G(K)$ the corresponding planar graph. Then the Jones polynomial $V_K(t)$ is equal to the Tutte polynomial $T_{G(K)}(-t, -t^{-1})$ up to a sign and factor a power of t.*

This is a remarkable observation from which it is now easy to deduce such facts as: the coefficients of the Jones polynomial of an alternating link alternate in sign (according to parity of degree). That is, *the Jones polynomial of an alternating link is an alternating polynomial.* The reader is referred to Thistlethwaite's paper for more details ([**T1**]).

VIII. From Graph Theory to Knot Theory

It is interesting to speculate about alternate realities. How could the bracket polynomial, and hence the Jones polynomial have emerged from graph theory?! One possible reconstruction is to suppose that graph theory had had in its possession our generalized Tutte polynomial for signed graphs. Since knots and links can be encoded into signed graphs, it would then have been possible to look for a specialization of this Tutte polynomial that gives an invariant.

To taste the flavor of this reconstruction we must first examine the graph theoretic versions of the Reidemeister moves. View FIGURE 22 for this. Because of

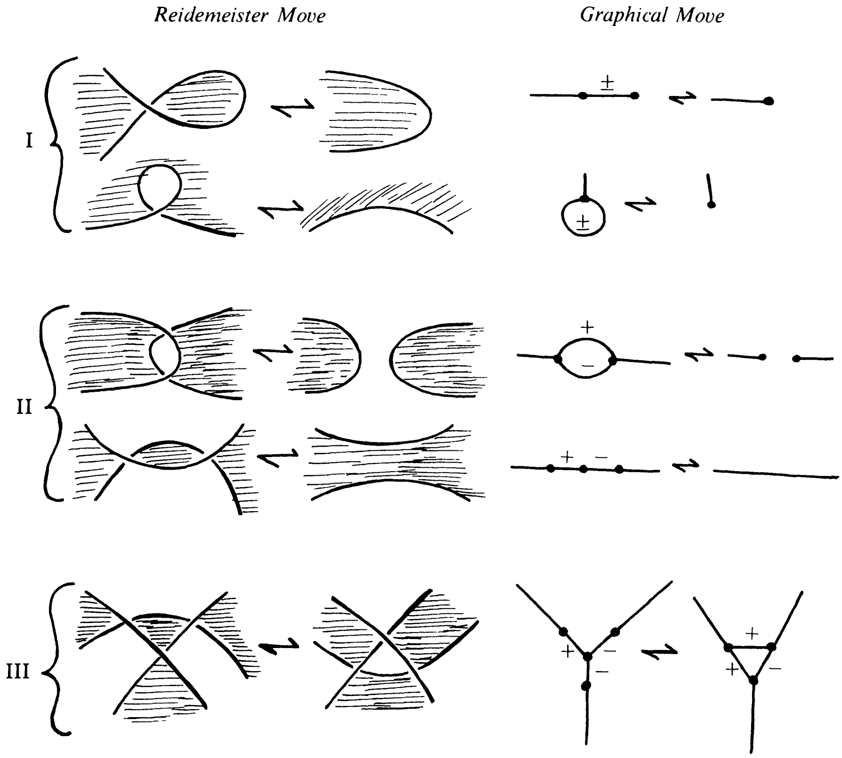


FIG. 22.

the translation to graphs via shaded regions, there are two versions of the type II move, and two versions of the type III move. Note that the two versions of the type II move are signed forms of deletion and contraction, while the type I move involves addition or removal of a branch or loop. These relationships could have been taken as a hint to try a Tutte polynomial for an invariant.

In any case, let's do this. We begin with the generalized Tutte polynomial for signed graphs, as explained in example 4 of the last section. This assigns a polynomial $T_G(A, B, x, y)$ to any signed graph G.

It is characterized by the rules: ($x = Ad + B$, $y = A + Bd$).

1) If G has only isthmuses and loops, then $T_G = x^{i_+ + l_-}y^{i_- + l_+}$ where i_+ is the number of positive isthmus, i_- is the number of negative isthmus, l_+ is the number of positive loops, l_- is the number of negative loops.

$$\text{2)}\quad T\;\overset{+}{—\bullet—\bullet—} = AT\;—\bullet— + BT\;—\bullet\quad\bullet—$$

$$T\;\overset{-}{—\bullet—\bullet—} = BT\;—\bullet— + AT\;—\bullet\quad\bullet—$$

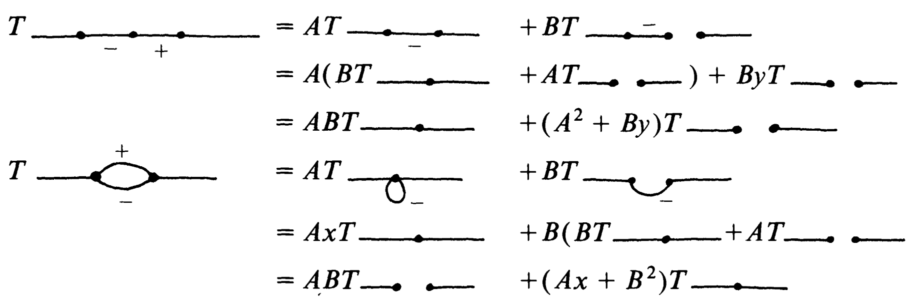
(the $\pm$ edge is not an isthmus.)

Let's investigate the behaviour of this polynomial under type II moves.

PROPOSITION 8.1. *In order for $T_G(A, B, x, y)$ to be invariant under type II moves it is necessary and sufficient that*

$$\begin{cases} B = A^{-1} \\ x = -A^{-3} \\ y = -A^{3}. \end{cases}$$

Proof.



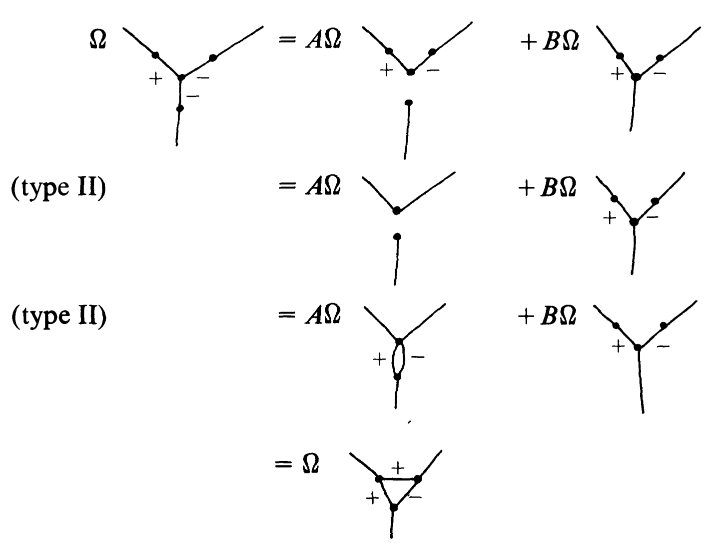
The rest of the proof follows from these identities.

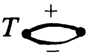
The rest of the story now proceeds just as for the bracket. With $B = A^{-1}$, $x = -A^{-3}$, $y = -A^3$ the polynomial T_G becomes an invariant of moves II *and* III for arbitrary graphs.

Definition 8.2. For any connected graph G, let $\Omega \in Z[A, A^{-1}]$ be the Laurent polynomial defined by $\Omega_G = T_G(A, A^{-1}, -A^{-3}, -A^3)$. By Proposition 8.1, Ω_G is an invariant of graphical move II. Type III invariance is free:

PROPOSITION 8.2. $\Omega_G = \Omega_{G'}$ *if G and G' are related by a type III move.*

Proof.



Remark. We want to be able to perform the type II move even when it disconnects the graph. Hence we need a formula for $T_{G \sqcup G'}$ where $\sqcup$ denotes disjoint union. Note that T$= Ax + By$. Define $T_{G \sqcup G'} = (Ax + By)T_G T_{G'}$. It is then easy to see that Ω has the right invariance properties for this move. Note that for $B = A^{-1}$, $x = -A^{-3}$, $y = -A^3$ we have $Ax + By = -A^{-2} - A^2 = d$, the corresponding bracket factor.

This section is just intended as a sketch of the graph-theoretical formulation. Note that Ω_G is an invariant of type II and type III moves for arbitrary (not necessarily planar) signed graphs. This is likely to be a useful extension of knot theory to arbitrary networks involving those transformations.

What about the type I move? ΩG multiplies by x or y under a type I move. In the abstract graph theoretic setting we do not have a direct analog of the twist number. Thus it remains to be seen whether ΩG can be normalized to form an invariant of all three move types.

To finish the translation we state the now obvious:

THEOREM 8.3. *Let G be a planar signed graph. Let $K(G)$ be the knot/link diagram corresponding to G. Then $\langle K(G) \rangle = \Omega_G$. The bracket polynomial for knots and links is a specialization of the generalized Tutte polynomial for signed graphs.*

Finally, we return to the generalized Tutte polynomial $T_G(A, B, x, y)$ for signed graphs G, and note that *it has a spanning tree expansion*. Given an ordering of the edges of G and a maximal tree $H \subset G$, define contributions from the edges of G as follows:

internally active,	$+ : x$
externally active,	$- : x$
internally active,	$- : y$
externally active,	$+ : y$
internally inactive,	$+ : A$
externally inactive,	$- : A$
internally inactive,	$- : B$
externally inactive,	$+ : B$

Let $G(H)$ denote the product of the contributions of the edges of G relative to activities for H. Then

$$T_G = \sum_H G(H),$$

where this summation extends over all maximal trees in G.

Technical Caveat. It is necessary and sufficient for T_G to be well-defined that $Ax + B^2 = A^2 + By$. (Note that we are in the category of signed graphs.) This certainly holds for the topological case. In the general case we can rephrase this

condition, as we did for the square bracket, by introducing a variable d and writing $x = A + Bd$, $y = Ad + B$. This shows that the square bracket is directly generalized to arbitrary signed graphs by $T_G(A, B, x, y) = T_G(A, B, d)$.

IX. The Knot Theory of Imbedded Graphs

It will not do to mention graphs without pointing out the now active extension of knot theory considering embeddings of arbitrary graphs in Euclidean three space. Here one would like to answer the usual ambient isotopy questions of knot theory in this larger context. In particular, we want polynomial (or other simple) invariants of graphs in space. The most general notion of ambient isotopy for graphs in space allows *topological vertices*. Strands coming into a topological vertex behave independently. Diagrammatically this means that moves are allowed (see FIGURE 23) that create arbitrary braiding at a vertex. At the vertex, any two adjacent strands can be given a twist. In general this notion of ambient isotopy is both fundamental and difficult. Nevertheless, progress is being made (see [S] and [W]) on the general classification.

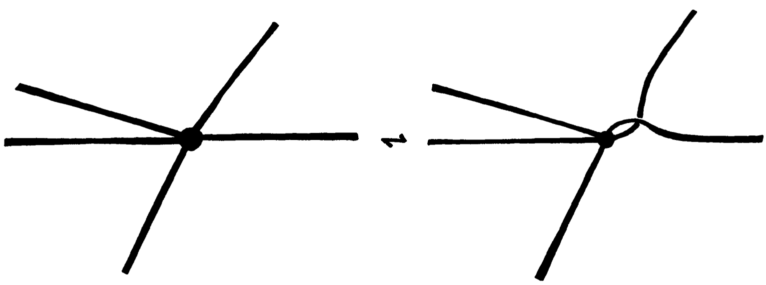


FIG. 23. A topological vertex move.

One may also consider a *rigid vertex*. Here the vertex is thought of as a rigid object with topological strands attached at specific sites. Then I note [**K8**] that it is possible to define some useful invariants in this case. A similar approach was seen independently by Ken Millett. Here is a sketch of my viewpoint about the rigid vertices. We shall restrict ourselves to 4-valent vertices as shown in FIGURE 24.
FIGURE 24 indicates the extra move-types that must be added to the list of Reidemeister moves in order to have a theory of rigid vertex equivalence. Note that the rigidity of the vertex forces double braiding when it is turned by 180°. I have denoted by III′ the analog of the type III move. The second move under IV (with a three strand twist) can be accomplished up to ambient isotopy by the first type IV. (We will not consider regular isotopy of graphs here.) My method for obtaining invariants of RV4-graphs (4-valent rigid vertex graphs) is to associate to such a graph G a collection of knots and links $\mathscr{L}(G)$ obtained as described below. This can be done for both oriented and nonoriented graphs. Here we consider only nonoriented graphs.

III′.

IV.

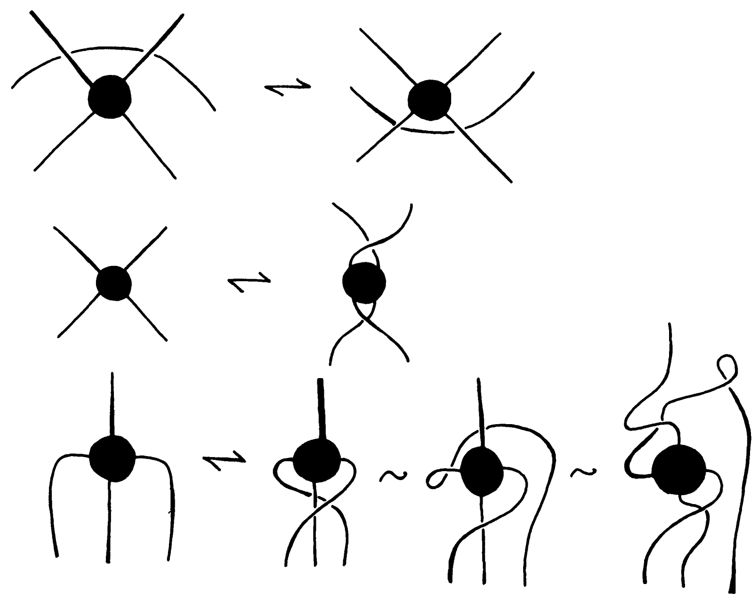
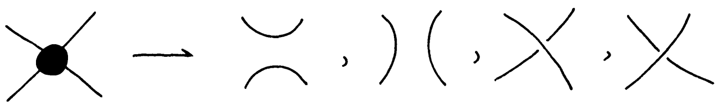
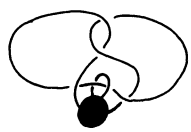


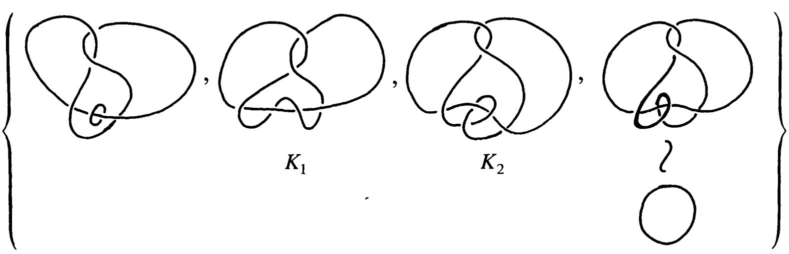
FIG. 24. Rigid vertex moves.

Definition 9.1. Let G be an RV4-graph. Let $\mathscr{L}(G)$ be the collection of knots and links obtained from G by choosing one replacement of each of the following types at each vertex of G:



Example.

$G =$  (J. Simon's graph.)

$\mathscr{L}(G) =$ 

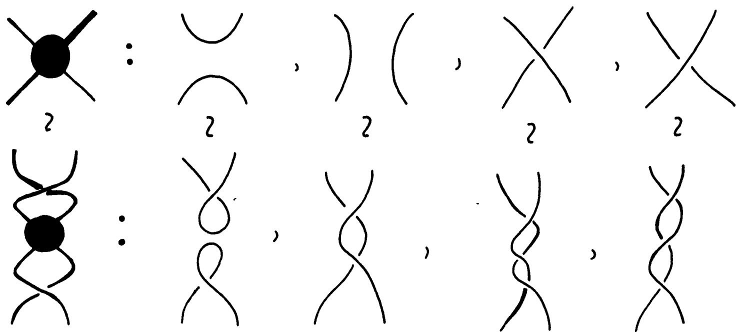
In general, if G has n rigid vertices, then $\mathscr{L}(G)$ will contain 4^n diagrams, some trivial, some ambient isotopic.

Definition 9.2. Let X be a collection of knots and links. Two such collections will be said to be *ambient isotopic* ($X \sim X'$) if every member of the first collection is ambient isotopic to some member of the second collection and vice versa.

The notation $\sim$ will be used both for RV4-equivalence of graphs and for ambient isotopy of knots and links.

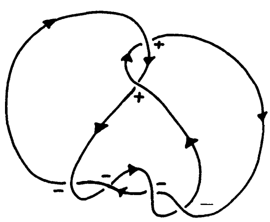
THEOREM 9.3. *Let G and G' be equivalent* RV4-*graphs in three-dimensional space. Then their associated link collections are ambient isotopic—$\mathscr{L}(G) \sim \mathscr{L}(G')$.*

Proof. Observe that the extra moves III′ and IV (FIGURE 24) preserve the elements of $\mathscr{L}(G)$ up to ambient isotopy. For example:

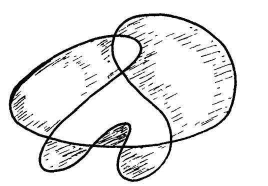


This completes the proof.

This is a very useful theorem for studying RV4-graphs. For example, we can immediately conclude that the graph G in the example above is not equivalent to its mirror image. For if this were so then the individual knots and links in $\mathscr{L}(G)$ (being distinct) would have to each be achiral. We can then check that this is not the case by using our results about the Jones polynomial for alternating knots. Recall that we have shown that if K is achiral then $3w(K) = W - B$ (after Theorem 3.1) where $w(K)$ is the twist number of K and W and B are the numbers of white and black regions in a shading where the A-channels are shaded. Choosing $K_1 \in \mathscr{L}(G)$ we find



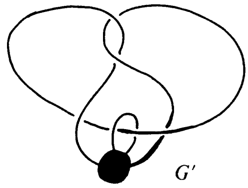
$$w(K_1) = 2 - 4 = -2$$
$$3w(K_1) = -6$$



$$W = 3$$
$$B = 5$$
$$W - B = -2$$

Since $-6 \neq -2$ we conclude that Simon's graph G is not equivalent (RV4) to its mirror image. (Wylbur Whitten has verified by other techniques that this graph is not topologically equivalent to its mirror image.)

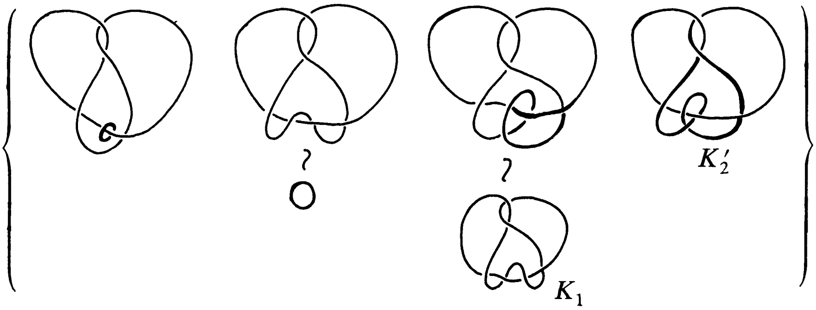
Example.



G'

Examination will reveal that this graph is obtained from the graph G of the previous example by reversing two crossings.

$\mathscr{L}(G') =$



We see that to prove G is not equivalent to G' it suffices to show that K_2' is not ambient isotopic to K_2. But we also know from section 3 that span$(K_2') < 4V$ where V is the number of crossings in this diagram, since K_2' is a non-alternating diagram. On the other hand, span$(K_2) = 4V$ ($V = 7$) since K_2 is an alternating diagram. Therefore, without further calculation we know that K_2 and K_2' are different and hence that G and G' are not RV4-equivalent graphs.

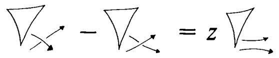
To further verify that G' is chiral requires a calculation of $\langle K_2' \rangle$. We omit this and assert that the calculation indeed shows that K_2' is chiral, and hence that G' is RV4-chiral.

These examples and our theorem relating equivalences of graphs with ambient isotopy for collections of knots and links show how there can be a good collaboration between problems of graph-imbedding and new invariants of knots and links such as the Jones polynomial. Since graphs can be used to model the configurations of molecules and other naturally-occuring networks, it is to be expected that there will be many fruitful applications of these ideas.

It is interesting to note that in the case of RV4-graphs there is a mixed-mechanical/topological model that is nevertheless susceptible to a topological analysis. This, in itself, is a good sign for applications where there will always be a mixture of topology and other structures.

X. Patterns and Speculations

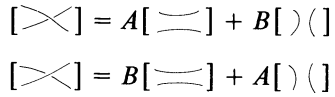
I always thought that the Conway identity



bore a striking resemblance to the exchange identities of quantum physics such as the Heisenberg form of the uncertainty principle $PQ - QP = \hbar i$. And that the crossing and its reverse were something like a complex number and its conjugate.

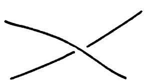
From the present vantage these speculations are not nonsense at all. They appear like hints about the remarkable connections that have subsequently appeared. The operator algebras that produced the Jones polynomial and generalizations were traditionally studied with quantum physics as the technical and inspirational source. The writing of the Conway identity into a Hecke algebra related to braids via $\sigma_i - \bar{\sigma}_i = z$ is a direct algebraic version of this sort of identity.

Another view of complex numbers lets us think of $[A, B]$ and $[B, A]$ as conjugates. (If $W = a + ib$ let $A = (1/2)(a + b)$, $B = (1/2)(a - b)$.) Then $[A, B] - [B, A] = (A - B)[1, -1] = (A - B)\sqrt{-1}$ has a formal resemblance to both the Conway identity and the Heisenberg formula. And in the bracket expansion



the reversal of a crossing appears as this form of conjugation.

Simple aspects of formalism lie at the root of these similarities. Beyond all the conceptual apparatus, the fundamental point is that an unoriented crossing



discriminates two regions out of four. All the rest, whether in braids or in diagram form builds on this distinction. Formulas like $PQ - QP = \hbar i$ rely on the left-right distinction along a line. And a formula such as [ ] = A[ ] + B[ ) ( ] relies on the discrimination of characters along a line that are identical after a rotation.

Thus we are initially mapping one simple order into another. These are designer's comments: for the possibility of distinguishing handedness of topological objects in three-dimensional space goes back through diagrams, symbols and distinctions to the possibility of finding handedness in the plane.

And the simplest forms of left and right in the plane are and . Thus the detection and discrimination of and (or and) in a formalism may be fundamental to its sensitivity to handedness. This is my personal explanation for favoring regular isotopy and detecting the curls in the diagrams.

There are more questions than you can shake a stick at. What deeper insight will unlock the really hidden secrets of these diagrams? Does the Jones polynomial, or its generalizations detect knottedness? Is there a further relationship to physics based on these diagrams (a lá Feynman diagrams or Penrose spin nets)? How can one understand the mirror image problem for knots and links completely? What is the relationship between these new techniques and the classical methods using homotopy theory, covering spaces and algebraic topology? What information about slice knots and knot concordance is in the new invariants? What is the next simple idea that will turn the subject upside down? How do you follow the hints?

REFERENCES

[A] J. W. Alexander, Topological invariants of knots and links, *Trans. Amer. Math. Soc.*, 30 (1923) 275–306.

[B] R. J. Baxter, Exactly Solved Models in Statistical Mechanics, Academic Press, 1982.

[BCW] William R. Bauer, F. H. C. Crick, and James H. White, Supercoiled DNA, *Sci. Am.*, 243 (1980) pp. 118–133.

[Bir] Joan S. Birman, Braids, links and mapping class groups, *Ann. of Math. Studies*, No. 82, 1976, Princeton University Press, Princeton, New Jersey.

[BW] Joan S. Birman and Hans Wenzel, Link polynomials and a new algebra (preprint).

[BLM] R. Brandt, W. B. R. Lickorish and K. C. Millett, A polynomial invariant for unoriented knots and links, *Math.* (1986) 503–13.

[Con] J. H. Conway, An enumeration of knots and links and some of their algebraic properties, Computational Problems in Abstract Algebra, Pergamon Press, New York, 1970, pp. 329–358.

[Co] Daryl Cooper, thesis, Warwick (1981).

[Cr] R. H. Crowell, Genus of alternating link types, *Ann. of Math.*, 69 (1959) 258–275.

[CF] R. H. Crowell and R. H. Fox, Introduction to Knot Theory, Blaisdell Publishing Company, 1963.

[F] F. B. Fuller, Decomposition of the linking number of a closed ribbon: A problem from molecular biology, *Proc. Natl. Acad. Sci.*, 75 (1978) 3557.

[G] C. Giller, A family of links and the Conway calculus, *Trans. Amer. Math. Soc.*, 270 (1982) 25–109.

[H] C. H. Ho, A new polynomial invariant for knots and links—preliminary report, *AMS Abstracts*, Vol. 6, #4, Issue 39 (1985), 300.

[HOMFLY] P. Freyd, D. Yetter, J. Hoste, W. Lickorish, K. Millett, and A. Ocneanu, A new polynomial invariant of knots and links, *Bull. Amer. Math. Soc.*, 12 (1985) 239–246.

[J1] V. F. R. Jones, A new knot polynomial and von Neumann Algebras, *Notices of AMS* (1985).

[J2] ______, A polynomial invariant for links via von Neumann algebras, *Bull. Amer. Math. Soc.*, 12 (1985) 103–112.

[J3] ______, Hecke algebra representations of braid groups and link polynomials (preprint).

[K1] L. H. Kauffman, The Conway polynomial, *Topology*, 20 (1980) 101–108.

[K2] ______, Formal Knot Theory, Princeton University Press, Mathematical Notes #30, 1983.

[K3] ______, On Knots, Princeton University Press, Annals Studies Number 15 (1987).

[K4] ______, An Invariant of Regular Isotopy (announcement—1985).

[K5] ______, State models and the Jones polynomial, *Topology*, Vol. 26, No. 3, 1987, pp. 395–407.

[K6] ______, Invariants of regular isotopy (to appear).

[K7] ______, Knots and Physics (in preparation).

[K8] ______, Invariants of graphs in three space (to appear).

[K9] ______, A Tutte Polynomial for Signed Graphs (to appear).

[Ki] M. Kidwell, On the degree of the Brandt-Lickorish-Millett polynomial of a Link (preprint).

[LM] W. B. R. Lickorish and K. Millett, A polynomial invariant of oriented links (to appear in *Topology*).

[L] W. B. R. Lickorish, A relationship between link polynomials, *Math. Proc. Cambridge Philos. Soc.* (to appear).

[M] H. Morton, Seifert circles and knot polynomials, *Math. Proc. Comb. Phil. Soc.*, 99 (1986) pp. 107–109.

[M1] K. Murasugi, On the genus of the alternating knot I, II, *J. Math. Soc. Japan*, 10 (1958) 94–105 and 235–248.

[M2] ______, Jones polynomials and classical conjectures in knot theory I and II, *Topology*, Vol. 26, No. 2, pp. 187–94.

[P] R. Penrose, Applications of negative dimensional tensors, in Combinatorial Mathematics and its Applications, Edited by D. J. A. Welsh, Academic Press, 1971.

[R] K. Reidemeister, *Knotentheorie*, Chelsea Publishing Company, New York, 1948.

[Ro] D. Rolfsen, Knots and Links, Publish or Perish Press, 1976.

[S] J. Simon, Topological chirality of certain molecules, *Topology*, 25 (1986) pp. 229–235.

[T1] M. Thistlethwaite, A spanning tree expansion of the Jones polynomial, *Topology*, Vol. 26, No. 3, pp. 297–309.

[T2] ______, Kauffman's polynomial and alternating links (to appear).

[Tu] W. T. Tutte, A contribution to the theory of chromatic polynomials, *Canad. J. Math.*, 6 (1953) 80–91.

[W] K. Wolcott, The knotting of theta curves and other graphs in S^3, (to appear in *Proceedings of* 1985 *Georgia Topology Conference*).

[Wh] J. White, Self-linking and Gauss integral in higher dimensions, *Amer. J. Math.*, July, Vol. XCI, (1969), 693–728.

[Wu] Y. Q. Wu, Jones polynomial and the crossing number of links (preprint).

[Y] D. Yetter, Markov algebras (preprint).

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

When Does a Polynomial over a Finite Field Permute the Elements of the Field?

RUDOLF LIDL*
Department of Mathematics, University of Tasmania, Hobart, Tasmania 7001, Australia

GARY L. MULLEN**
Department of Mathematics, Pennsylvania State University, University Park, PA 16802

If $f(x) = a_0 + a_1x + \cdots + a_nx^n$ is a polynomial of degree n over the finite field F_q of order $q = p^\delta$ where p is a prime and $\delta \geqslant 1$, does the associated polynomial function f: $c \to f(c)$ from F_q into F_q permute the elements of F_q, i.e., is f a 1-1 map of F_q onto itself? If f is a permutation of F_q then the polynomial $f(x)$ is called a **permutation polynomial** (PP) of F_q. Very little is known concerning which polynomials are PPs, despite the attention of numerous authors (see the notes to Chapter 7 of [**13**]). Very few good algorithms exist to test whether a given polynomial is a PP. We describe several applications of permutations which indicate why a study of permutations is of interest, and we list, without proof, the known classes of PPs and the known criteria. We indicate a number of open problems and make some conjectures.

Recently, permutations of finite fields have become of considerable interest in the construction of cryptographic systems for the secure transmission of data. See, for example [**12**, Chapter 9] and [**10**]–[**11**]. Let M be a message (an element of F_q) which is to be sent securely from A to B. If $P(x)$ is a permutation of F_q, then A sends to B the field element $N = P(M)$. Since $P(x)$ is a bijection, B can obtain the original message M by calculating $P^{-1}(N) = P^{-1}(P(M)) = M$. In order to be useful in a cryptographic system, $P(x)$ must have several additional properties; see problem P3. Cryptographic systems of a different type, based on permutations of finite fields, were considered by Levine and Brawley in [**10**].

*Research was supported by the Australian Research Grants Committee, grant F8415183.

**This author would like to thank the mathematics department at the University of Tasmania for its hospitality during a recent visit when discussions which led to this paper took place.

Permutations are also useful in several combinatorial applications. The reader is referred to Theorems 9.67 and 9.83 of [**13**] as well as [**4**] and [**19**]–[**20**] for illustrations thereof.

We now list the major known classes of PPs. If $f(x)$ is a PP of F_q, then so is $f_1(x) = af(x+b)+c$ for all $a \neq 0, b, c \in F_q$ and we say that $f_1(x)$ is in **normalized form** if a, b, c are chosen so that $f_1(x)$ is monic, $f_1(0) = 0$, and (provided the characteristic p does not divide the degree n) the coefficient of x^{n-1} is 0.

(1) In [**5**, p. 63] Dickson lists all normalized permutations of degree at most 5.

(2) x^k permutes F_q if and only if $(k, q-1) = 1$.

(3) If $a \in F_q$, then the Dickson polynomial

$$g_k(x,a) = \sum_{j=0}^{[k/2]} \frac{k}{k-j}\binom{k-j}{j}(-a)^j x^{k-2j}$$

permutes F_q if and only if $(k, q^2-1) = 1$ [**13**, Theorem 7.16].

(4) If F_{q^r} is an extension of F_q of degree r, then $L(x) = \Sigma_{s=0}^{r-1}\alpha_s x^{q^s}$ with $\alpha_s \in F_{q^r}$ is a linear operator on F_{q^r} over F_q and it permutes F_{q^r} if and only if $\det(\alpha_{i-j}^{q^j}) \neq 0$ $(i, j = 0, 1, \ldots, r-1)$. If each $\alpha_s \in F_q$ then $L(x)$ permutes F_{q^r} if and only if $(\Sigma_{s=0}^{r-1}\alpha_s x^s, x^r - 1) = 1$ [**13**, pp. 362, 390].

(5) If $r > 1$ is prime to $q-1$, s divides $q-1$, and $g(x^s)$ has no nonzero root in F_q, where $g(x) \in F_q[x]$, then $x^r(g(x^s))^{(q-1)/s}$ permutes F_q [**13**, Theorem 7.10].

(6) Several other specific classes of PPs have been characterized [**13**, Chapter 7], e.g., $x^{(q+m-1)/m} + ax$ where m divides $q-1$ and $x^r(x^d - a)^{(p^n-1)/d}$, where $d | (p^n - 1)$.

This list is not exhaustive, but it does contain all known major classes: so we don't have very many types of PPs.

Suppose we wish to determine whether $f(x) = \Sigma_{i=0}^n a_i x^i$ permutes F_q. We may assume that $n \leqslant q-1$ since every function from a finite field F_q to itself can be represented by a polynomial of degree $< q$ [**13**, p. 369]. If q is small, one could proceed by calculating $f(a)$ for all $a \in F_q$ and checking whether the q values $f(a)$ are indeed distinct. However since this involves $O(qn)$ F_q operations for a polynomial f of degree n, this becomes prohibitive if q is large.

The first, and in some ways most useful, criterion was proved by Hermite [**8**] for q prime, and by Dickson [**6**] for general q.

THEOREM. *$f(x)$ permutes F_q if and only if*

(1) f has exactly one root in F_q and

(2) For each integer t with $1 \leqslant t \leqslant q-2$ with $t \not\equiv 0 \pmod p$, the reduction of $[f(x)]^t \bmod (x^q - x)$ has degree $\leqslant q-2$.

COROLLARY. *If $f(x)$ is a PP of F_q of degree $n > 1$, then $n \nmid (q-1)$.*

London and Ziegler [**14**], Raussnitz [**21**], and Mollin and Small [**16**] investigated criteria for $f(x)$ to be a permutation in terms of the coefficients of $f(x)$. Two sufficient conditions can be found in Carlitz and Lutz [**2**].

We now list a number of open problems and conjectures involving PPs of finite fields.

P1. Find an algorithm of lower complexity than $O(qn)$ to test whether a given polynomial is a PP of F_q.

P2. Find new classes of PPs of F_q.

P3. Find new classes of permutations $P(x)$ that are useful cryptographically. Hence $P(x)$ should have a simple form so that if M is a message, then $N = P(M)$ which is sent from A to B can be easily computed. Also $P(x)$ must have the property that without some secret information (the key) that only A and B know, $P^{-1}(x)$ will be hard or impossible to get, so that an unauthorized receiver cannot calculate $P^{-1}(N)$. At the same time with knowledge of the key, $P^{-1}(x)$ is easily obtained by B so that $P^{-1}(N) = M$ can be recovered by B.

P4. The Chebyshev polynomial of the second kind of degree k is defined by

$$f_k(x) = \sum_{i=0}^{[k/2]} \binom{k-i}{i} (-1)^i x^{k-2i}.$$

It can be shown that $f_k(x) = xf_{k-1}(x) - f_{k-2}(x)$ for $k \geqslant 2$. For odd q determine necessary and sufficient conditions on k and q so that $f_k(x)$ permutes F_q. In **[15]** Matthews showed that for odd q, if $k+1 \equiv \pm 2(\operatorname{mod} m)$ for $m = p$, $(q-1)/2$, and $(q+1)/2$, then $f_k(x)$ permutes F_q, and in fact then, $f_k(-a) = -f_k(a)$ and $f_k(a) = \pm a$ for all $a \in F_q$. Based upon computer evidence, we conjecture that if q is prime, then the above conditions are also necessary. Moreover if the characteristic $p > 5$ it appears that the conditions are both necessary and sufficient for $f_k(x)$ to permute F_q. For $p = 3$ and 5, there are examples of $q > p$ and of k such that $f_k(x)$ permutes F_q but k does not satisfy the above set of congruences.

P5. Extend Dickson's list [**5**, Table 7.1] of normalized PPs to higher degrees.

P6. Let $N_d = N_d(q)$ denote the number of PPs of degree d over F_q. We have the trivial boundary conditions $N_1 = q(q-1)$, $N_d = 0$ if $d|(q-1)$, and $\Sigma N_d = q!$ where the sum is over all $1 \leqslant d < q-1$ with $d \nmid (q-1)$. Find N_d. Some partial results in this direction were given in Wells **[24]**.

P7. Suppose $f(x)$ is a PP of degree p over F_q. Dickson **[5]** conjectured that $f(x)$ is reducible to the normalized PP $x(x^d - \alpha)^{(p-1)/d}$ where $d|(p-1)$ and $\alpha \neq \beta^d$ for any $\beta \in F_q$, having earlier proved this for $p = 3$, 5, and 7 in his thesis **[6]**. Settle this conjecture for all odd primes p.

P8. Settle the conjecture of Chowla and Zassenhaus [3] that if p is a sufficiently large prime and $f(x)$ of degree $\geqslant 2$ permutes F_p, then $f(x) + ax$ with $0 < a < p$ is not a PP of F_p. For several partial results, see **[20]** and **[17]**.

P9. In an invited address before the MAA in 1966, Carlitz conjectured that for each even positive integer k, there is a constant C_k such that for each finite field of odd order $q > C_k$, there does not exist a PP of degree k over F_q. Lausch and Nöbauer [**9**, p. 202] proved the conjecture for $k = 2^m$ and for $k = 6$ and 10 the conjecture has been proven by Dickson **[6]** and Hayes [7], respectively. Recently Wan Daqing **[22]** settled the cases $k = 12$ and 14. A major result would be a proof of the Carlitz conjecture for arbitrary even k.

Many similar questions can be considered for polynomials in several variables over F_q, see e.g. [**13**, Sect. 7.5] and for residue class rings of integers, see for example **[9]** and [**18**], where there are numerous other references.

Acknowledgment. The authors would like to thank the referee and the editor for a number of suggestions that improved the final version of the paper.

REFERENCES

1. L. Carlitz, Permutations in a finite field, *Acta Sci. Math. Szeged*, 24 (1963) 196–203.
2. L. Carlitz and J. A. Lutz, A characterization of permutation polynomials over a finite field, this MONTHLY, 85 (1978) 746–748.
3. S. Chowla and H. Zassenhaus, Some conjectures concerning finite fields, *Norske Vid. Selsk. Forh.* (Trondheim), 41 (1968) 34–35.
4. S. D. Cohen and M. J. Ganley, Some classes of translation planes, *Quart. J. Math., Oxford Ser.*, (2) 35 (1984) 101–113.
5. L. E. Dickson, Linear Groups with an Exposition of the Galois Field Theory, Teubner, Leipzig, 1901; Dover, New York, 1958.
6. L. E. Dickson, The analytic representation of substitutions on a power of a prime number of letters with a discussion of the linear group, *Ann. of Math.*, 11 (1897) 65–120, 161–183.
7. D. R. Hayes, A geometric approach to permutation polynomials over a finite field, *Duke Math. J.*, 34 (1967) 293–305.
8. C. Hermite, Sur les fonctions de sept lettres, *C. R. Acad. Sci. Paris*, 57 (1863), 750–757; Oeuvres, vol. 2, Gauthier-Villars, Paris, 1908, 280–288.
9. H. Lausch and W. Nöbauer, Algebra of Polynomials, North-Holland, Amsterdam, 1973.
10. J. Levine and J. V. Brawley, Some cryptographic applications of permutation polynomials, *Cryptologia*, 1 (1977) 76–92.
11. R. Lidl and W. B. Müller, A note on polynomials and functions in algebraic cryptography., *Ars Combin.*, 17A (1984) 223–229.
12. R. Lidl and H. Niederreiter, Introduction to Finite Fields and their Applications, Cambridge University Press, Cambridge, 1986.
13. R. Lidl and H. Niederreiter, Finite Fields, Encyclopedia Math. Appl., Vol. 20, Addison-Wesley, Reading, Mass 1983 (now distributed by Cambridge Univ. Press).
14. D. London and Z. Ziegler, Functions over the residue field modulo a prime, *J. Austral. Math. Soc., Ser. A* 7 (1967) 410–416.
15. R. W. Matthews, Permutation Polynomials in One and Several Variables, Ph.D. Dissertation, University of Tasmania, 1982.
16. R. A. Mollin and C. Small, On permutation polynomials over finite fields, *Internat. J. Math. Math. Sci.*, 10 (1987) 535–544.
17. G. L. Mullen and H. Niederreiter, Dickson polynomials over finite fields and complete mappings, *Canad. Math. Bull.*, 30 (1987) 19–27.
18. W. Narkiewicz, Uniform Distribution of Sequences of Integers in Residue Classes, Lecture Notes in Math., Vol. 1087, Springer-Verlag, New York, 1984.
19. H. Niederreiter and K. H. Robinson, Bol loops of order pq, *Math. Proc. Cambridge Philos. Soc.*, 89 (1981) 241–256.
20. H. Niederreiter and K. H. Robinson, Complete mappings of finite fields, *J. Austral. Math. Soc.*, Ser A 33 (1982) 197–212.
21. G. Raussnitz, *Math. Naturw. Ber. Ungarn*, 1 (1882/83) 266–278.
22. Wan Daqing, On a conjecture of Carlitz, *J. Austral. Math. Soc., Ser. A* (43) 1987, 375–84.
23. C. Wells, Groups of permutation polynomials, *Monatsh. Math.*, 71 (1967) 248–262.
24. C. Wells, The degrees of permutation polynomials over finite fields, *J. Combinatorial Theory*, 7 (1969) 49–55.

Notes

EDITED BY DENNIS DETURCK, RICHARD LIBERA, AND ANITA E. SOLOW

Homomorphisms from $\mathbf{Z_m}[\mathbf{i}]$ into $\mathbf{Z_n}[\mathbf{i}]$ and $\mathbf{Z_m}[\rho]$ into $\mathbf{Z_n}[\rho]$, Where $i^2 + 1 = 0$ and $\rho^2 + \rho + 1 = 0$

JOSEPH A. GALLIAN
Department of Mathematics and Statistics, University of Minnesota, Duluth 55812

DOUGLAS S. JUNGREIS
Department of Mathematics, Harvard University, Cambridge, MA 02138

In a recent note [1] it was shown that the number of group homomorphisms from $\mathbf{Z_m}$ into $\mathbf{Z_n}$ is $\gcd(m, n)$ and the number of ring homomorphisms from $\mathbf{Z_m}$ into $\mathbf{Z_n}$ is $2^{\omega(n)-\omega(n/\gcd(m,n))}$ where $\omega(u)$ denotes the number of distinct prime factors of the integer u. In this note we solve the analogous problems for $\mathbf{Z}[\mathbf{i}]$ and $\mathbf{Z}[\rho]$, where $i^2 + 1 = 0$ and $\rho^2 + \rho + 1 = 0$.

We begin with the groups $\mathbf{Z}[\mathbf{i}]$ and $\mathbf{Z}[\rho]$. The number of group homomorphisms from $\mathbf{Z_m}[\mathbf{i}]$ into $\mathbf{Z_n}[\mathbf{i}]$ and from $\mathbf{Z_m}[\rho]$ into $\mathbf{Z_n}[\rho]$ is $(\gcd(m, n))^4$. This follows directly from the solution in $\mathbf{Z}$, since the groups $\mathbf{Z_k}[\mathbf{i}]$ and $\mathbf{Z_k}[\rho]$ are each isomorphic to $\mathbf{Z_k} \oplus \mathbf{Z_k}$. In the ring case, the solutions and their derivations are more complicated.

For convenience, let $G = \mathbf{Z}[i]$ denote the ring of Gaussian integers and for any Gaussian integer α let G_α denote the factor ring $G/\langle\alpha\rangle$. We seek explicit formulae for the number of homomorphisms from G into G_n and from G_m into G_n.

THEOREM 1. *The number of ring homomorphisms from G into G_n is $c_n \cdot 3^{\omega(n)}$, where $\omega(n)$ is the number of distinct prime factors of n in G and where $c_n = 1$ if $4 \nmid n$, $c_n = 5/3$ if $4|n$ but $8 \nmid n$, $c_n = 3$ if $8|n$.*

Proof. First observe that a homomorphism from G into G_n is completely determined by its action on 1 and i. Since $1^2 = 1, 1 \cdot i = i$, and $i^2 = -1$ in G, the images of 1 and i, call them a and b respectively, satisfy $a^2 = a$, $a \cdot b = b$, and $b^2 = -a$ in G_n. Let $n = p_1^{k_1} \cdots p_r^{k_r}$ be the prime-power decomposition of n in G. Since, by the Chinese Remainder Theorem, G_n is naturally isomorphic to $G_{p_1^{k_1}} \oplus G_{p_2^{k_2}} \oplus \cdots \oplus G_{p_r^{k_r}}$, a homomorphism from G into G_n induces a homomorphism from G into $G_{p_j^{k_j}}$ for each j. In the direct sum, let a correspond to $(a_1, a_2, \dots, a_r)$ and b correspond to $(b_1, b_2, \dots, b_r)$.

Then $a_j^2 \equiv a_j \bmod p_j^{k_j}$, or equivalently $p_j^{k_j} | a_j(a_j - 1)$. Since G is a unique factorization domain and a_j and $a_j - 1$ are relatively prime either $p_j^{k_j} | a_j$ or $p_j^{k_j} | (a_j - 1)$. Thus either $a_j = 0$ or $a_j = 1$. If $a_j = 0$, then $b_j = a_j b_j = 0$. If $a_j = 1$, then $b_j^2 = -a_j = -1$, so $(b_j + i)(b_j - i) = 0$, and $p_j^{k_j} | (b_j + i)(b_j - i)$. Suppose $p_j \neq 1 + i$. Then, since the only prime divisor of $2i$ is $1 + i$ (up to associates), we

know p_j cannot divide both factors $b_j + i$ and $b_j - i$. Thus $p_j^{k_j} | b_j + i$ or $p_j^{k_j} | b_j - i$ and $b_j = \pm i$.

When $p_j = 1 + i$, complications arise because $(1+i)^2 = 2i$, so that $1 + i$ can divide both factors. Since n is an integer and $|1 + i| = \sqrt{2}$, we see that k_j, the exponent of $p_j = 1 + i$, must be even, say $k_j = 2t$. We consider three cases. Since $G_{p_j^{k_j}}$ is isomorphic to G_2 or G_4 respectively when $k_j = 2$ or 4, direct calculations show that there are respectively 2 or 4 possibilities for b_j. In the case $k_j \geqslant 6$, we must observe that $(1+i)^2$ divides one factor if and only if it divides the other, but $(1+i)^3$ cannot divide both factors. Thus $(1+i)^{k_j - 2}$ must divide one of the factors. Because $(1+i)^{k_j-2} = ((1+i)^2)^{t-1} = 2^{t-1}$ (up to associates) and $\{2^{t-1}(r+si)\}$ has 4 elements in $G_{2^t} \approx G_{p_j^{k_j}}$ we see that $(1+i)^{k_j-2}$ divides 4 elements in $G_{p_j^{k_j}}$. This gives 8 possible values for b_j.

Thus far we have determined necessary conditions for the existence of a homomorphism from G into G_n. Conversely, if $(a_1, a_2, \ldots, a_r)$ and $(b_1, b_2, \ldots, b_r)$ are any elements of the direct sum which satisfy the above conditions, then there is a homomorphism from G into G_n. So, the number of homomorphisms is as claimed. ■

To determine the number of homomorphisms from G_m into G_n we simply observe that the same conditions as above are necessary, and, furthermore, when $a_j = 1$ in $G_{p_j^{k_j}}$, we must have $ma_j = m = 0$ in $G_{p_j^{k_j}}$ (since $m \cdot 1 = 0$ in G_m). So, whenever $p_j^{k_j}$ does not divide m only the trivial homomorphism to that component of the direct product is possible. Thus, there is one homomorphism instead of 3, 5 or 9 as before and we must reduce the exponent of 3 or 5 accordingly. This proves the following.

THEOREM 2. *The number of ring homomorphisms from G_m into G_n is $c_n \cdot 3^{\omega(n) - \omega(n/\gcd(m,n))}$ where $\omega(k)$ is the number of distinct prime factors of k in G and $c_n = 1$ if either $4 \nmid n$ or $2 | (n/\gcd(m,n))$, $c_n = 5/3$ if $2 \nmid (n/\gcd(m,n))$ and $4|n$ but $8 \nmid n$ and $c_n = 3$ if $8|n$ and $2 \nmid (n/\gcd(m,n))$.*

We remark that for any integer k it is easy to determine $\omega(k)$ in G from the integer prime-power decomposition of k. Indeed, $\omega(k)$ in G is the number of distinct integer prime divisors of k plus the number of integer prime divisors of k that are congruent to 1 modulo 4. This follows from the fact that 2 and prime integers that are congruent to 3 modulo 4 have only one Gaussian prime divisor (up to associates) while prime integers that are congruent to 1 modulo 4 have two.

Another familiar ring of algebraic integers is $\mathbf{Z}[\rho]$ where $\rho^2 + \rho + 1 = 0$ (see Chapter 12 of [2] for a thorough treatment of $\mathbf{Z}[\rho]$). This ring was introduced by Eisenstein and Jacobi in their work on cubic reciprocity. The arguments used in Theorems 1 and 2 can be adapted to yield the following two results.

THEOREM 3. *The number of ring homomorphisms from $\mathbf{Z}[\rho]$ into $\mathbf{Z}_n[\rho]$ is $c_n \cdot 3^{\omega(n)}$, where $\omega(n)$ is the number of distinct prime factors of n in $\mathbf{Z}[\rho]$ and where $c_n = 1$ if $3 \nmid n$, $c_n = 4/3$ if $3|n$ but $9 \nmid n$, and $c_n = 7/3$ if $9|n$.*

Proof. Let the images of 1 and ρ be a and b, $G = \mathbf{Z}[\rho]$ and $G_\alpha = G/\langle\alpha\rangle$. Then $a^2 = a$, $a \cdot b = b$ and $b^2 + b + a = 0$ in G_n. Let a_j and b_j be as in the proof of

Theorem 1. Then, as before, $a_j = 0$ or 1 and the former implies $b_j = 0$. If $a_j = 1$, then $b_j^2 + b_j + 1 = 0$ so $(b_j - \rho)(b_j - \rho^2) = 0$ and $p_j^{k_j} | (b_j - \rho)(b_j - \rho^2)$. Suppose $p_j \neq 1 - \rho$. Then since the only prime divisor of the difference of these factors is $1 - \rho$, (up to associates), we know p_j cannot divide both factors. Thus, $b_j = \rho$ or ρ^2.

When $p_j = 1 - \rho$, complications arise because $1 - \rho$ can divide both factors. Since n is an integer and $|1 - \rho| = \sqrt{3}$, we see that k_j, the exponent of $p_j = 1 - \rho$, must be even. We consider two cases. Since $G_{p_j^{k_j}}$ is isomorphic to G_3 when $k_j = 2$, direct calculations show that there are 3 possibilities for b_j. In the case $k_j \geq 4$, $(1 - \rho)^{k_j - 1}$ divides one factor. Because $\mathbf{Z}[\rho]/\langle 1 - \rho\rangle$ has 3 elements, $(1 - \rho)^{k_j - 1}$ divides 3 elements in $G_{p_j^{k_j}}$. This gives 7 possible values for b_j. The results now follows as in Theorem 1. ■

THEOREM 4. *The number of ring homomorphisms from* $\mathbf{Z}_m[\rho]$ *into* $\mathbf{Z}_n[\rho]$ *is* $c_n \cdot 3^{\omega(n) - \omega(n/\gcd(m,n))}$ *where* $\omega(k)$ *is the number of distinct prime factors of* k *in* $\mathbf{Z}[\rho]$ *and* $c_n = 1$ *if either* $3 \nmid n$ *or* $3 | (n/\gcd(m, n))$, $c_n = 4/3$ *if* $3 \nmid (n/\gcd(m,n))$ *and* $3 | n$ *but* $9 \nmid n$, *and* $c_n = 7/3$ *if* $3 \nmid (n/\gcd(m, n))$ *and* $9 | n$.

Proof. The analysis is identical to that given in the proof of Theorem 2. ■

The authors are indebted to I. H. Rose and David Witte for detecting errors in our original versions of Theorems 3 and 4.

This work was done while the second author was visiting the University of Minnesota, Duluth. Both authors were supported by the NSF (Grant number DMS 8407498).

REFERENCES

1. Joseph A. Gallian and James Van Buskirk, The Number of Homomorphisms from $\mathbf{Z}_m$ into $\mathbf{Z}_n$, *American Math. Monthly*, 91 (1984) 196–197.
2. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, Oxford University Press, 4th ed., 1960.

An Elementary Proof That a Bounded a.e. Continuous Function is Riemann Integrable

MICHAEL W. BOTSKO
Department of Mathematics, Saint Vincent College, Latrobe, PA 15650

In [**1**, p. 142], the following Theorem is given.

THEOREM 1. *If f is continuous on $[a, b]$, then it is Riemann integrable there.*

For the proof the function F: $[a, b] \to \mathbf{R}$ is defined such that

$$F(x) = \overline{\int_a^x} f(t)\, dt - \underline{\int_a^x} f(t)\, dt \quad \text{for } a < x \leq b \quad \text{and} \quad F(a) = 0.$$

It is then noted that $F'(x) = 0$ for all x in $[a, b]$, which implies that F is constant on $[a, b]$. From this fact it is easy to see that the upper and lower integrals are equal.

It seems to be desirable to have a proof of this type, on the advanced calculus level, for the more general situation where the function f is bounded and continuous a.e. on $[a, b]$. It is the purpose of this paper to present such a proof. We first recall the definition of a set of measure zero.

DEFINITION. A set E of real numbers has *measure zero* if for each $\varepsilon > 0$ there is a finite or infinite sequence $\{I_n\}$ of open intervals covering E and satisfying $\Sigma_n |I_n| \leqslant \varepsilon$ where $|I_n|$ is the length of I_n.

We begin by giving an advanced calculus proof of the following well-known result.

LEMMA 1. *If $f\colon [a, b] \to \mathbf{R}$ satisfies a Lipschitz condition and $f'(x) = 0$ except on a set of measure zero, then f is a constant function.*

Proof. Let ε and δ be arbitrary positive numbers. Since f satisfies a Lipschitz condition there exists $M > 0$ such that $|f(x) - f(y)| \leqslant M|x - y|$ for all x, y in $[a, b]$. Let $D = \{x : x \in [a, b] \text{ and } f'(x) = 0\}$ so that $[a, b] \setminus D$ has measure zero. For $\delta/M > 0$ there is a sequence $\{I_n\}$ of open intervals such that $[a, b] \setminus D \subseteq \cup I_n$ and $\Sigma_n |I_n| \leqslant \delta/M$. Now for $I = [c, d]$ any closed subinterval of $[a, b]$, we will say that I is a type 1 interval if $|f(d) - f(c)| \leqslant \varepsilon(d - c)$ and a type 2 interval if there exists $J \in \{I_n\}$ such that $I \subseteq J$. We now define the set

$S = \{x : x \in (a, b]$ and there exists $P_x = \{x_0, x_1, x_2, \ldots, x_n\}$ a partition of $[a, x]$ such that for each k, $[x_{k-1}, x_k]$ is either a type 1 or type 2 interval$\}$.

We first note that $S \neq \varnothing$. If $a \in D$, then $f'(a) = 0$ and there exists $x > a$ such that $|f(x) - f(a)| \leqslant \varepsilon(x - a)$, which says that $x \in S$. (Note that a partition may contain only two points.) On the other hand, if $a \notin D$ then $a \in J$ an interval in $\{I_n\}$ and, since J is open, there exists $x > a$ such that $[a, x] \subseteq J$ and again $x \in S$.

Next we show that if $x \in S$, then

$$|f(x) - f(a)| \leqslant \delta + \varepsilon(b - a).$$

To see this take a partition $P_x = \{x_0, x_1, x_2, \ldots, x_n\}$ of $[a, x]$ such that each $[x_{k-1}, x_k]$ is either a type 1 or a type 2 interval. Clearly

$$f(x) - f(a) = \sum_k f(x_k) - f(x_{k-1})$$

so that

$$\begin{aligned} |f(x) - f(a)| &\leqslant \sum_k |f(x_k) - f(x_{k-1})| \\ &= \sum_{(1)} |f(x_k) - f(x_{k-1})| + \sum_{(2)} |f(x_k) - f(x_{k-1})|. \end{aligned}$$

(The first summation is taken over type-1 intervals of P_x and the second over type-2

intervals of P_x.)

$$\leqslant \sum_{(1)} \varepsilon(x_k - x_{k-1}) + \sum_{(2)} M(x_k - x_{k-1})$$

$$\leqslant \varepsilon(b-a) + M\sum_{(2)}(x_k - x_{k-1})$$

$$\leqslant \varepsilon(b-a) + \delta.$$

The last inequality holds since $\Sigma_n\ |I_n| \leqslant \delta/M$ and the subintervals involved are nonoverlapping type-2 intervals. Thus $x \in S$ implies that

$$|f(x) - f(a)| \leqslant \delta + \varepsilon(b-a).$$

Now let $c =$ the least upper bound of S. It follows at once that $a < c \leqslant b$ and since f is continuous, $|f(c) - f(a)| \leqslant \delta + \varepsilon(b-a)$. We need to show that $c = b$. Suppose $c < b$. If $c \in D$ then $f'(c) = 0$. Thus there exists $r > 0$ such that $[c-r, c+r] \subseteq (a, b)$ and

$$|f(x) - f(c)| \leqslant \varepsilon|x - c| \quad \text{if } x \in [c-r, c+r].$$

Since $c - r$ is not an upper bound of S, there exists $x \in S$ such that $c - r < x \leqslant c$. Thus there is $P_x = \{x_0, x_1, x_2, \dots, x_n\}$, a partition of $[a, x]$ such that each subinterval is either type 1 or type 2. Clearly $P = \{x_0, x_1, x_2, \dots, x_n, c + r\}$ is a partition of $[a, c+r]$ where each subinterval is of type 1 or type 2. Thus $c + r \in S$ which is a contradiction. If $c \notin D$ then $c \in J$ for some J in $\{I_n\}$, and there exists $r > 0$ such that $[c-r, c+r] \subseteq J$. Again there is an $x \in S$ such that $c - r < x \leqslant c$. By an argument similar to the previous case $c + r \in S$, again a contradiction. Thus $c = b$ and

$$|f(b) - f(a)| \leqslant \delta + \varepsilon(b-a).$$

Since δ and ε are arbitrary positive numbers, $f(b) = f(a)$ and it clearly follows that $f(x) = f(a)$ for all x in $[a, b]$, which completes the proof.

We can now easily accomplish the main task of this paper.

THEOREM 2. *If $f:[a,b] \to \mathbf{R}$ is bounded, and continuous except on a set of measure zero, then f is Riemann integrable on $[a, b]$.*

Proof. Let

$$F(x) = \overline{\int_a^x} f(t)\,dt - \underline{\int_a^x} f(t)\,dt \quad \text{if } a < x \leqslant b \quad \text{and } F(a) = 0.$$

Since f is bounded, F clearly satisfies a Lipschitz condition on $[a, b]$. (To see this use Lemmas 5.2e and 5.2g in [**1**, pp. 137–138].) Also by Lemma 5.2h in [**1**, p. 139], $F'(x) = f(x) - f(x) = 0$ for each x where f is continuous. Thus $F'(x) = 0$ except on a set of measure zero and by Lemma 1, F is a constant function. Thus

$$0 = F(a) = F(b) = \overline{\int_a^b} f(t)\,dt - \underline{\int_a^b} f(t)\,dt.$$

Thus f is Riemann integrable on $[a, b]$ since the upper and lower integrals are equal.

REFERENCE

1. Watson Fulks, Advanced Calculus, Wiley, New York, 1978.

L'Hôpital's Rule in a Poisson Derivation

WILLIAM P. COOKE
Department of Mathematics and Physical Sciences, West Texas State University, Canyon, TX 79016

Elementary derivations of the Poisson distribution have two basic forms, which we might call direct and indirect. Given assumptions of independence and negligibility of probabilities for more than one arrival in a small time interval, the assumption that the probability of an arrival in a small time interval is *proportional* to the length of the interval implies a direct derivation (cf. [**1**]). When we just assume that the arrival probability *depends* on the length of the interval, then we must first develop the exponential density of interarrival times—the indirect approach (cf. [**4**]). Here L'Hôpital's rule (see [**3**, p. 121]) will be used in the indirect approach to yield a simple alternative to the usual derivation as shown in Wagner [**4**].

The direct derivation is more suitable for a first course in probability, where discrete distributions are usually discussed early. In a course on queuing theory or operations research, however, this alternate indirect derivation could simplify the teaching of the topic.

From the assumptions in Wagner [**4**] the exponential density of interarrival times follows:

$$f(t) = \begin{cases} \lambda \exp(-\lambda t), & \lambda > 0,\ t > 0 \\ 0, & \text{elsewhere.} \end{cases} \tag{1}$$

Define $P_n(t)$ as

$$P_n(t) = \text{Prob}[n \text{ arrivals in the interval } (0, t)].$$

Our independence and negligibility assumptions give

$$P_n(t+h) = P_n(t)P_0(h) + P_{n-1}(t)P_1(h), \qquad n = 1, 2, \ldots, \tag{2}$$

if $h > 0$ is regarded as being very small. Subtracting $P_n(t)$ from both sides of (2), dividing by h, and taking the limit gives the derivative

$$P_n'(t) = \lim_{h \to 0} \left[\{P_n(t)[P_0(h) - 1] + P_{n-1}(t)P_1(h)\}/h\right], \qquad n = 1, 2, \ldots. \tag{3}$$

From (1) we obtain

$$P_0(h) = \int_h^\infty \lambda \exp(-\lambda t)\, dt = \exp(-\lambda h), \tag{4}$$

and our negligibility assumption gives

$$P_1(h) = 1 - P_0(h) = 1 - \exp(-\lambda h).$$

Then (3) becomes

$$P_n'(t) = P_n(t)\lim_{h\to 0}\{[\exp(-\lambda h) - 1]/h\} \\ + P_{n-1}(t)\lim_{h\to 0}\{[1 - \exp(-\lambda h)]/h\}, \qquad n=1,2,\dots. \tag{5}$$

Each of the two limits in (5) has the indeterminate form 0/0, but one application of L'Hôpital's rule to each limit produces

$$P_n'(t) = -\lambda P_n(t) + \lambda P_{n-1}(t), \qquad n = 1,2,\dots.$$

Then we can use either mathematical induction or the probability generating function (cf. [2]), along with (4), to show that

$$P_n(t) = [(\lambda t)^n \exp(-\lambda t)]/n!, \qquad n = 0,1,2,\dots,$$

the Poisson distribution with parameter λt.

REFERENCES

1. J. E. Freund, Mathematical Statistics, 2nd ed., p. 85, Prentice-Hall, 1971.
2. H. A. Taha, Operations Research, 3rd ed., pp. 634–635, Macmillan, 1982.
3. A. E. Taylor, Advanced Calculus, Blaisdell, 1955.
4. H. M. Wagner, Principles of Operations Research, pp. 846–847, Prentice-Hall, 1969.

Problems that Teach the Obvious but Difficult

NEAL KOBLITZ
Department of Mathematics, University of Washington, Seattle, WA 98195

Most of our students are supposed to be learning calculus and other subjects in order to be able to recognize and apply the concepts in contexts which may be far removed from the settings that typify textbook problems. The transition is likely to present obstacles for them. Some simple matters which we take for granted and generally ignore should, in my opinion, be a major consideration in our design of exercises. I would like to illustrate by discussing four of my favorite calculus word problems.

Problem 1. *Find the centroid of Nevada.*

Here we make a slight simplification, giving Nevada a small piece of Arizona, so that it becomes a perfect trapezoid. We also assume that 1° is a constant distance throughout Nevada. The student is given the latitude and longitude of the four corners of Nevada: 42°N, 120°W; 42°N, 114°W; 39°N, 120°W; 35°N, 114°W. This problem at first stumps many students because of the absence of any equations and the unfamiliar form in which the coordinates of the corners are given. The first step in solving the problem is to decide how to set up the xy-axes. One can choose the

conventional picture, with the point $(0°, 0°)$ (the equatorial point on the Greenwich meridian) as origin and east as the positive x-direction. But in that case, the student is confronted with fairly large negative numbers for the coordinates and is almost guaranteed to make an error in computing the equation of the slanted side. The shrewder move is to choose the origin to be one of the two corners along the slanted side, and define x and y to be degrees longitude and latitude relative to that reference point. Then the student can easily find the equations of the top and bottom of Nevada, and the rest of the solution is routine even for the average student.

Problem 2. *When crude oil flows from a well, water is frequently mixed with it in an emulsion. To remove the water the crude is piped to a device called a heater-treater, which is simply a large tank in which the oil is warmed and the water is allowed to settle out. Operating experience in a particular oil field indicates that the concentration C of water in the treater's output can be modeled by the following equation in a neighborhood of the usual operating point of* $135°F$ *and a* 2-*hour holding time*:

$$C = 0.04 - 0.0032h^2 - 10^{-6}t^2 + 4.1 \cdot 10^{-5}ht,$$

where h is the holding time in hours and t is the operating temperature in degrees F. (*a*) *Because of random fluctuations in the well's flow rate, the holding time actually varies slightly around* 2 *hours. Suppose you are given a simple control device that can change the tank temperature proportionally to the measured change in holding time. What constant of proportionality would best compensate for small holding time fluctuations and keep the water concentration as constant as possible?* (*b*) *Now as field equipment ages, its maximum operating settings are generally decreased. Find the equation of the line that best approximates the way in which the holding time would have to be increased as the maximum temperature rating falls slightly below the usual operating temperature.*

All this problem asks for is the derivative (dt/dh in part (a) and dh/dt in part (b)), which is easily found by implicit differentiation. But most students find this exercise quite difficult, because they have trouble distilling the essence of the question from all the words. They are thrown off because the derivative and the tangent line are described using "practical" rather than mathematical terms. Unfortunately, most students are so locked into the textbook terminology that they do not necessarily recognize the concept of derivative when it occurs "in nature."

Problem 2 was given to me by an engineering student, who said it was a problem he had been asked to solve while working one summer for an oil company.

Problem 3. *Your friend is at the top of a building which you know to be exactly* 100 *feet high. She throws a stone downward at v_0 ft/sec, and you time how long it takes to hit the ground. Neglect air resistance, so that the distance* (*in feet*) *fallen by the stone in t seconds is given by the formula* $s = 16t^2 + v_0t$. *If you time* 2 *seconds for the stone to fall to the ground, and if your stopwatch is accurate only to within* ± 0.1 *sec, at what downward speed did your friend throw the stone? Include an error estimate in your answer. Use the tangent line approximation.*

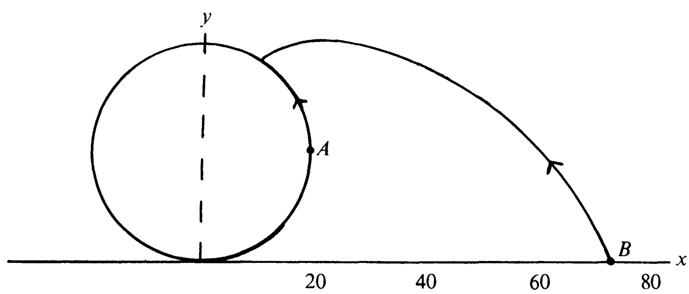
Here a student who has just mastered the derivation of the falling body formula is almost certain to attempt to treat t and s as the variables and v_0 as a constant. How can v_0 *not* be a constant? Once that psychological barrier is surmounted, and the student realizes that in this particular problem s is constant ($= 100$) and v_0 is the dependent variable, the error in the value for v_0 is seen to be simply $(\pm 0.1)(dv_0/dt)$.

Here is a problem of a similar sort for more advanced students, who have studied functions of two variables.

Problem 3A. *If an object falls vertically from rest with air resistance proportional to velocity, then its velocity at time t is given by the formula* $v(t) = \frac{g}{k}(1 - e^{-kt})$. *Here* $g = -9.8\ m/sec^2$ *is known exactly, and* k, *the constant of proportionality between the air resistance and the velocity, is measured experimentally with some error* Δk. (a) *Derive a formula for the height* $s(t)$ *in terms of the initial height* s_0 *and* g, k, t. (b) *Suppose you measure that it takes* 2.0 *seconds for the object to hit the ground, with an error of* Δt *in this measurement. Your wind tunnel experiments give a value of* 0.1 *for* k, *with an error of* Δk. *Find a simple formula for* s_0 *in terms of* Δk *and* Δt. *Use the tangent plane approximation.*

The formula for $v(t)$ in this problem is obtained by solving the differential equation $dv/dt = g - kv$ with initial condition $v(0) = 0$; then one obtains $s(t) = s_0 + \frac{g}{k^2}(kt + e^{-kt} - 1)$ by integration. After working with t, v, and s as the variables and g, k, s_0 as the constants, the student must then shift gears in order to answer the question in part (b). That is, she or he must set $s(t)$ equal to zero and regard s_0 as the dependent variable, which is a function of the two independent variables k and t. The answer: $18.4 - 11.8\Delta k + 17.8\Delta t$.

Problem 4. *You are standing on the ground at point B (see diagram), a distance of* 75 *feet from the bottom of a ferris wheel that is* 20 *feet in radius. Your arm is at the same level as the bottom of the ferris wheel. Your friend is on the ferris wheel, which makes one revolution (counterclockwise) every* 12 *seconds. At the instant when she is at point A you throw a ball to her at* 60 ft/sec *at an angle of* 60° *above the horizontal. Take* $g = -32\ ft/sec^2$, *and neglect air resistance. Find the closest distance the ball gets to your friend, using Newton's method to obtain an answer which is accurate to within* 1/2 *foot.*



In this problem the student must first organize all the information that is presented and divide the problem into manageable steps. These steps might be: (a) ignoring the ball for a minute, find the friend's angular velocity in rad/sec, and find parametric equations for the friend's position at time t; (b) leaving the friend for a minute to look at the ball, resolve the ball's initial velocity into x- and y-components (noticing that the horizontal motion is in the negative x-direction), and find parametric equations for the ball's position at time t; (c) express the distance between the ball and the friend in terms of t; (d) find the derivative of the expression in part (c) and set it equal to zero (a slightly simpler approach would be to minimize the *square* of the distance); (e) guess an approximate time t_0 when the ball is fairly close to the friend (take $t_0 = 2$), and use Newton's method with one iteration to find a better approximate value t_1 when the ball is closest to the friend; (f) use part (c) to find the distance at time t_1; (g) check that taking a second iteration in Newton's method does not appreciably change the value you get for the closest distance. The answer turns out to be: 1.9 ft when $t = 2.22$ sec.

Most textbook problems involve only one technique. Problem 4, on the other hand, employs almost all of the topics in a standard first-semester calculus course: use of trig functions, derivatives of trig functions, the chain rule, parametrized curves (falling body problems, circular motion), distance formula, max/min, and Newton's method. In this way it demands that the student bring together a variety of techniques into a multi-step thought process.

These four problems illustrate some aspects of the transition from mechanical "pure-math" problem solving to "real-world" problem solving: the value of choosing a convenient setup of the coordinate system; the need to recognize basic concepts, such as the derivative, when they are disguised in non-mathematical words; the importance of determining at the very beginning which letters stand for constants and which letters stand for variables; the usefulness of breaking up a problem into bite-sized chunks. This might all seem obvious to us; but unless students are drilled with problems that are designed to illustrate these features, they are likely to have trouble using the material from their math courses in their subsequent work in other fields.

PROBLEMS AND SOLUTIONS

EDITED BY PAUL. T. BATEMAN, HAROLD G. DIAMOND, KENNETH B. STOLARSKY (ADVANCED PROBLEMS), AND DOUGLAS B. WEST (ELEMENTARY PROBLEMS)

*For instructions about submitting **proposed** problems for publication in this department see the inside front cover. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the address given on the inside front cover. Two copies suffice.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

For instructions about submitting solutions of Problems, which should be mailed before July 30, 1988, see the inside front cover. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label if an acknowledgment is desired.

E 3253. *Proposed by Sydney Bulman-Fleming and Edward T. H. Wang, Wilfrid Laurier University, Waterloo, Canada.*

(a) Let n and t be natural numbers with n odd and $t \leq n$. Suppose x_i is an integer not divisible by n for $i = 1, 2, \dots, t$. Determine the maximum possible number of pairs (i, j) with $1 \leq i < j \leq t$ such that $x_i + x_j \equiv 0 \pmod{n}$.

(b) Let k, n, and t be natural numbers with n relatively prime to k and $t \leq n$. Suppose x_i is an integer not divisible by n for $i = 1, 2, \dots, t$. Determine the maximum possible number of k-tuples $(i_1, i_2, \dots, i_k)$ with $1 \leq i_1 < i_2 < \cdots < i_k \leq t$ such that $x_{i_1} + x_{i_2} + \cdots + x_{i_k} \equiv 0 \pmod{n}$.

E 3254. *Proposed by David Singmaster, Polytechnic of the South Bank, London, England.*

If a_0 and c are given positive integers, define a sequence $\{a_n\}$ by letting a_{n+1} be the positive integer obtained by reversing the digits in the ordinary decimal expansion of $a_n + c$ $(n = 0, 1, 2, \dots)$. For example, if $a_0 = 18$ and $c = 2$, the sequence begins 18, 2, 4, 6, 8, 1, 3, 5, 7, 9, 11, 31, 33, 53, 55, 75, 77, 97, ...

(a) If $c = 1$, show that the sequence $\{a_n\}$ is eventually periodic and that all such sequences have the same ultimate cycle.

(b) If $c = 2$ and $1 \leq a_0 \leq 99$, show that the sequence $\{a_n\}$ is eventually periodic.

(c)* If $c = 2$, show that the sequence $\{a_n\}$ is eventually periodic for any choice of a_0.

E 3255. *Proposed by Paul Erdös, Hungarian Academy of Sciences.*

What is the minimum number of edges in a connected n-vertex graph such that every edge belongs to a triangle?

E 3256. *Proposed by John Isbell, SUNY at Buffalo.*

(a) Let T be the set of triangles in the *plane* whose vertices have integral coordinates and whose sides have integral lengths. Certain isosceles triangles in T can be constructed by fitting together two congruent right triangles in T, e.g., the isosceles triangle with vertices $(-12, -9)$, $(12, 9)$, $(-12, 16)$ arises in this way. Are there any other isosceles triangles in T?

(b) Consider the set V of triangles in 3-*space* whose vertices have integral coordinates. Does V contain any equilateral triangles with integral side-length?

E 3257. *Proposed by I. A. Sakmar, Istanbul, Turkey.*

Let P, Q, R be the new vertices of equilateral triangles constructed outwardly on the edges of a given triangle ABC. (Cf. H. S. M. Coxeter, *Introduction to Geometry*, New York, 1961, p. 22.)

(a) Show that any triangle PQR which can be obtained in this way arises from a unique triangle ABC, and give a construction for recovering triangle ABC from triangle PQR.

(b) Show that not every triangle PQR can be so obtained.

E 3258. *Proposed by Nicolae Gonciulea, Traian College, Drobeta Turnu Severin, Romania.*

Prove that

$$\sum_{j=0}^{n} \binom{n}{j} 2^{n-j} \binom{j}{\lfloor j/2 \rfloor} = \binom{2n+1}{n},$$

where $\lfloor x \rfloor$ is the largest integer not exceeding x.

SOLUTIONS OF ELEMENTARY PROBLEMS

A Quantifier Question

E 2984 [1983, 54]. *Proposed by Robert Maddux, Iowa State University.*

Characterize the logically valid sentences of the form

$$Q_1 x_1 \cdots Q_n x_n R x_1 \cdots x_n \rightarrow Q_1' x_{i_1} \cdots Q_n' x_{i_n} R x_1 \cdots x_n, \qquad (*)$$

where Q_i is $\forall$ or $\exists$, $i_1, \ldots, i_n$ is a permutation of $1, \ldots, n$, and R is a relation.

Editorial Comment. The third line of the statement of the problem should read: "where each Q_i and each Q'_j is $\forall$ or $\exists$, $i_1, \dots, i_n$ is a permutation of $1, \dots, n$, and R is a relation symbol."

Solution by an anonymous referee, based on a partial solution by William M. Lambert, Jr. We show that the validity of (*) is equivalent to the conjunction of conditions $(*)_1$ and $(*)_2$:

If $\exists x_k$ occurs on the left in (*), it also occurs on the right. $(*)_1$

One cannot go from $\cdots \forall x_j \cdots \exists x_k \cdots R$ to $(*)_2$

$$\cdots \exists x_k \cdots \forall x_j \cdots R.$$

$(*) \Rightarrow (*)_1$: Proof by contraposition. Here (*) is of the form:

$$\cdots \exists x_k \cdots Rx_1 \cdots x_n \to \cdots \forall x_k \cdots Rx_1 \cdots x_n.$$

Consider the interpretation given by the relation $\tilde{R} \subseteq \omega^n$: $\tilde{R}(m_1, \dots, m_n)$ iff $m_k = 0$. Obviously the antecedent in (*) is true and the conclusion false, whence (*) is not valid.

$(*) \Rightarrow (*)_2$: Proof by contraposition. Here (*) is of the form:

$$\cdots \forall x_j \cdots \exists x_k \cdots R \to \cdots \exists x_k \cdots \forall x_j \cdots R.$$

Consider the interpretation given by the relation $\tilde{R} \subseteq \omega$: $\tilde{R}(m_1, \dots, m_n)$ iff $m_j = m_k$. Again, the antecedent in (*) is true under this interpretation while the succedent is false.

$(*)_1 \,\&\, (*)_2 \Rightarrow (*)$: The passage from $Q_1x_1 \cdots Q_nx_nR$ to $Q'_1x_{i_1} \cdots Q'_nx_{i_n}R$ is the result of two types of operations:

(a) change some $\forall$ to $\exists$,

(b) permute some (Qx) and $(Q'x')$.

Clearly, we can perform the (a) operations one at a time. Moreover, since every permutation can be written as a product of transpositions, we can perform only one of these operations at a time. Accordingly, there is a sequence

$$Q_1x_1 \cdots Q_nx_nR = \Phi_0, \qquad \Phi_1, \dots, \Phi_t = Q'_1x_{i_1} \cdots Q'_nx_{i_n}R,$$

where each Φ_{k+1} arises from Φ_k by means of one of the following operations:

(a) change $\forall$ to $\exists$: $\cdots \forall x_k \cdots R \to \cdots \exists x_k \cdots R$

(b¹) transpose $\forall$'s: $\cdots \forall x_k \forall x_j \cdots R \to \cdots \forall x_j \forall x_k \cdots R$

(b²) transpose $\exists$'s: $\cdots \exists x_k \exists x_j \cdots R \to \cdots \exists x_j \exists x_k \cdots R$

(b³) kill uniformity: $\cdots \exists x_k \forall x_j \cdots R \to \cdots \forall x_j \exists x_k \cdots R.$

[N.B. $(*)_1$ rules out changing $\exists$ to $\forall$; $(*)_2$ rules out introducing uniformity, i.e., changing $\forall\exists$ to $\exists\forall$.] But now it is clear that each operation preserves truth: (b¹) & (b²) do not change meaning and (a) & (b³) merely weaken the assertion. Hence

$$\Phi_0 \to \Phi_1, \qquad \Phi_1 \to \Phi_2, \dots, \Phi_{t-1} \to \Phi_t$$

are all valid, whence $\Phi_0 \to \Phi_t$, i.e., (*), is valid. Q.E.D.

Enveloping Circle

E 3084 [1985, 287]. *Proposed by Paris Pamfilos, University of Crete, Greece.*

Given a family of circles in the plane all of which pass through a common point and no two of which have the same diameter, show that there is another circle enveloping all the circles of the family if and only if there is a straight line containing all the intersection points of the common tangents of any two circles of the family.

Solution by O. P. Lossers, Eindhoven University of Technology, The Netherlands. Let the common point be denoted by P. We now apply inversion with respect to P and use the fact that such an inversion permutes the set of all circles and all lines in the plane conformally.

The modified problem reads as follows:

Given a point P and a family F of lines in the plane at distinct positive distances from P, show that the members of F have a common tangent circle, Γ say, if and only if there exists a circle, Δ say, passing through P such that the common tangent circles of any pair of lines l, $m \in F$ that pass through P share a second point $Q \in \Delta$.

Given the circle Γ with centre M and two tangent lines l and m with intersection point S, say, symmetry arguments immediately reveal that the common tangent circles of l and m containing P have in common the point Q that is obtained by reflecting P with respect to MS. Because M is fixed the orbit of Q is a circle Δ that has M as its centre, too.

Given the circle Δ with centre N and two lines l and m such that their common tangent circles containing P have point $Q \in \Delta$ in common, again by symmetry arguments we see that l and m have the same distance from N, in other words, l and m are tangent lines of the same circle concentric with Δ. This circle is called Γ.

Also solved by A. Bondesen (Denmark), J. Dou (Spain), J. Fukuta (Japan), L. Kuipers (Switzerland) and P. Szűsz, and the proposer. Both Bondesen and Fukuta established the (slightly) stronger fact that the assumption "all of which pass through a common point..." could be replaced by "all of which are tangent to a common circle, externally or internally... ."

Symmetric Positive Definite Matrices

E 3131 [1986, 132]. *Proposed by William N. Anderson, Jr., Fairleigh Dickinson University, and George E. Trapp, West Virginia University.*

Let A be a partitioned real matrix with submatrices A_{ij}, $i, j = 1, 2, \ldots, k$. Let B be the $k \times k$ matrix with elements b_{ij} given as follows: b_{ij} is the algebraic sum of all of the elements in the A_{ij} submatrix. Show that if A is symmetric positive definite, then B is symmetric positive definite.

Editorial note: Many solvers noted that the problem was incorrect as stated, unless by a partitioned matrix one understands that the diagonal submatrices A_{ii}

are square. For example,

$$A = \left[\begin{array}{cc|c} 1 & 0 & 1 \\ \hline 0 & 1 & 0 \\ 1 & 0 & 2 \end{array}\right]$$

is both symmetric and positive definite, but

$$B = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$$

has neither property.

Solution by A. A. Jagers, Technische Hogeschool Twente, Enschede, The Netherlands. Assume that the rows and columns of A are partitioned the same way. Let $C = [c_{ij}]$ be the point-block incidence matrix of this partition; that is, let $c_{ij} = 1$ if i belongs to the jth block of the partition and let $c_{ij} = 0$ otherwise. Then C has rank k and $B = C^TAC$, where C^T denotes the transpose of C. Hence B is symmetric positive definite.

Also solved by S. W. Bent, J. H. Bolstad, A. J. Bosch (The Netherlands), Chico Problem Group, T. Jager, O. Krafft (West Germany), O. P. Lossers (The Netherlands), J. M. Monier (France), D. Neuenschwander (Switzerland), B. Poonen, G. S. Rogers, A. Tissier (France), W. P. Wardlaw, P. Y. Wu (China), and the proposer.

Arithmetic-Geometric Mean

E 3142 [1986, 299]. *Proposed by Zhang Zaiming, Yuxi Teachers College, China.*

Prove the following refinement of the arithmetic-geometric mean inequality:

$$\sqrt{ab} = G(a,b) \leqslant L(A,b) \leqslant I(a,b) \leqslant A(a,b) = (a+b)/2 \quad \text{for } 0 < a \leqslant b,$$

where the logarithmic mean is

$$L(a,b) = \begin{cases} (b-a)/(\log_e b - \log_e a) & \text{if } 0 < a < b \\ a & \text{if } 0 < a = b \end{cases}$$

and the identric mean is

$$I(a,b) = \begin{cases} \dfrac{1}{e}\left(b^b/a^a\right)^{1/(b-a)} & \text{if } 0 < a < b \\ a & \text{if } 0 < a = b \end{cases}.$$

(M. E. Mays, Functions which parametrize means, this MONTHLY, 90 (1983) 677–683.)

Editorial comment. There was a misprint in the original statement of the problem, namely, (a^a/b^b) was erroneously given in place of (b^b/a^a) in the definition of $I(a,b)$.

Solution by Ricardo Perez Marco (student), Barcelona, Spain. When $a = b$ we have a string of equalities. For $0 < a < b$, let $r = b/a > 1$. Dividing the desired

inequalities by a, it suffices to prove the strict inequalities

$$\sqrt{r} < (r-1)/\ln r < \frac{1}{e} r^{r/(r-1)} < (r+1)/2 \quad \text{for } r > 1.$$

1) For the first inequality, it suffices to show $g(r) = \ln r - \sqrt{r} + 1/\sqrt{r} < 0$. Observe that $g'(r) = -(1-\sqrt{r})^2/(2r\sqrt{r}) < 0$ and $g(1) = 0$.

2) For the second inequality, it suffices to show

$$h(r) = \frac{1}{r-1} r^{r/(r-1)} \ln r > e.$$

Since $\lim_{r\to 1^+} h(r) = e$, we need only show $h' > 0$ for $r > 1$. In this range $h > 0$, so we simply compute

$$h'/h = \frac{d}{dr}\ln h = \frac{1}{r\ln r}\left[1 - r\left(\frac{\ln r}{r-1}\right)^2\right].$$

By the first inequality, the bracketed quantity is positive, as desired.

3) For the third inequality, it suffices to show

$$f(r) = r^{r/(r-1)}/(r+1) < e/2.$$

Taking a similar approach, we note $\lim_{r\to 1^+} f(r) = e/2$ and want $f' < 0$ for $r > 1$. In this range $f > 0$, so again we compute

$$f'/f = \frac{d}{dr}\ln f.$$

The result of this is $k(r)/(r-1)^2$, where

$$k(r) = 2\frac{r-1}{r+1} - \ln r.$$

Now we have what we want, since $k(1) = 0$ and $k'(r) = -(r-1)^2/\{r(r+1)^2\} < 0$ for $r > 1$.

Editorial comment. The Chico Problem Group, I. Merényi, and R. E. Pfiefer employed the inequalities

$$\left(\frac{1}{r-1}\int_1^r \frac{dx}{x}\right)^{-1} \leqslant \exp\left(\frac{1}{r-1}\int_1^r \ln x\,dx\right) \leqslant \frac{1}{r-1}\int_1^r x\,dx.$$

D. M. Bloom gave bounds for the spread in the inequalities. Many solvers noted that solutions or generalizations have appeared in many places, including the following:

H. Alzer, Über einen Wert, der zwischen dem geometrischen und dem arithmetischen Mittel zweier Zahlen liegt, *Elemente Math.*, 40 (1985) 22–24.

H. Alzer, Ungleichungen für $(e/a)^a(b/e)^b$, *Elemente Math.*, 40 (1985) 120–123.

B. C. Carlson, A hypergeometric mean value, *Proc. Am. Math. Soc.*, 16 (1965) 759–766.

B. C. Carlson and M. D. Tobey, A property of the hypergeometric mean value, *Proc. Am. Math. Soc.*, 19 (1968) 255–262.

E. B. Leach and M. C. Sholander, Extended mean values, this MONTHLY, 85 (1978) 84–90.
K. B. Stolarsky, Generalizations of the logarithmic mean, *Mathematics Magazine*, 48 (1975) 87–92.

Also solved by S. Arslanajić (Yugoslavia), D. M. Bloom, P. Bracken (Canada), J. L. Brenner, B. Brunson, Chico Problem Group, W. Herget, W. Janous (Austria), S. V. Keny, P. K. Khalili, L. Kuipers (Switzerland), I. Merényi (Romania), J. E. Pečarić (Yugoslavia), A. Pedersen (Denmark), R. E. Pfiefer, D. B. Tyler, M. Vowe (Switzerland), C. Wildhagen (Netherlands), and the proposer.

ADVANCED PROBLEMS

6568. *Proposed by Itshak Borosh and Douglas Hensley, Texas A & M University*

A lattice point $(s, t) \in N \times N$ is called primitive if the greatest common divisor of s and t is 1. It is known that the density of primitive lattice points is $6/\pi^2 = 0.6079\ldots$ (Hardy and Wright, *Theory of Numbers*, Theorem 459). Show that the density of lattice points $(a, b) \in N \times N$ for which each of the eight neighbors $(a \pm 1, b)$, $(a, b \pm 1)$, $(a \pm 1, b \pm 1)$ is primitive is

$$\frac{1}{4}\prod(1 - 8/p^2) = 0.012\ldots,$$

where the product is taken over all odd primes p.

6569. *Proposed by Wilson Castrellon and Samuel Camargo, Universidad de Los Andes, Bogotá, Colombia*

Show that for every infinite cardinal λ there exists a cardinal $\kappa \geq \lambda$, a set A of cardinality κ, and a family F of subsets of A such that $\operatorname{card}(F) = 2^\kappa$ and $X \cap Y$ is finite for all $X, Y \in F$, $X \neq Y$.

SOLUTIONS OF ADVANCED PROBLEMS

A Legendre Polynomial Identity

6517 [1986, 305]. *Proposed by Alexandru Lupaş, Facultatea de mecanică, Sibiu, Romania.*

If P_n is the sequence of Legendre polynomials, i.e.,

$$P_n(x) = \frac{1}{2^n n!}\left(\frac{d}{dx}\right)^n [(x^2 - 1)^n],$$

show that

$$[P_n(x)]^2 - P_{n-1}(x)P_{n+1}(x) = \frac{1}{\pi n(n+1)}\int_{-1}^{1} \frac{1 - P_n[x^2 + t(1 - x^2)]}{1 - t}\frac{dt}{\sqrt{1 - t^2}}.$$

Editorial Note. The original statement of this problem contained the erroneous factor $2h_n$ on the right side of the integral identity, where h_n is defined below.

Solution by the proposer. Let

$$D_n(x) = P_n^2(x) - P_{n-1}(x)P_{n+1}(x)$$

be Turán's Legendre polynomial determinant. By an equality attributed to L. Moser and M. Wyman (see D. S. Mitrinović, *Specijalne Funkcije*, Beograd, 1978, formula (2.2.67)) we have

$$D_n(x) = \frac{1-x^2}{n(n+1)} \sum_{k=0}^{n-1} \frac{1}{k+1} \sum_{j=0}^{k} (2j+1)P_j^2(x).$$

Apply to this the summation by parts formula

$$\sum_{k=0}^{n-1} A_k B_k = \sum_{k=0}^{n-2} (A_k - A_{k+1}) \sum_{j=0}^{k} B_j + A_{n-1} \sum_{j=0}^{n-1} B_j$$

with $B_k = 1/(k+1)$ to obtain

$$D_n(x) = \frac{1-x^2}{n(n+1)} \sum_{j=0}^{n-1} (2j+1)P_j^2(x)(h_n - h_j),$$

where $h_0 = 0$ and

$$h_n = \sum_{k=1}^{n} \frac{1}{k} \quad \text{for } n \geq 1.$$

Next, the product formula for Legendre polynomials (see A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions*, vol. 1, McGraw-Hill, New York, 1953, formula 3.15(19)) is

$$P_j(x)P_j(y) = \frac{1}{\pi}\int_{-1}^{1} P_j\left(xy + t\sqrt{1-x^2}\sqrt{1-y^2}\right)\frac{dt}{\sqrt{1-t^2}}.$$

Therefore,

$$P_j^2(x) = \frac{1}{\pi}\int_{-1}^{1} P_j(x^2 + t(1-x^2))\frac{dt}{\sqrt{1-t^2}}.$$

Now use the Christoffel-Darboux Summation formula (see M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover, New York, 1965, Chapter 22, formula 22.12.1) with $y = 1$ to obtain

$$\frac{P_{n-1}(x) - P_n(x)}{1-x} = \sum_{j=0}^{n-1} \frac{1}{n}(2j+1)P_j(x).$$

On the left is a telescoping summand, so it is easy to see that a further summation

followed by a change in the order of summation on the right leads to

$$\frac{1-P_n(x)}{1-x}=\sum_{j=0}^{n-1}(h_n-h_j)(2j+1)P_j(x).$$

Hence,

$$\frac{1-P_n(x^2+t(1-x^2))}{1-t}=(1-x^2)\sum_{j=0}^{n-1}(h_n-h_j)(2j+1)P_j(x^2+t(1-x^2)).$$

Upon multiplying both sides by $(1-t^2)^{-1/2}$, integrating with respect to t, and using the formulas derived above for $P_j^2(x)$ and $D_n(x)$, we obtain

$$\frac{1}{\pi}\int_{-1}^{1}\frac{1-P_n(x^2+t(1-x^2))}{1-t}\frac{dt}{(1-t^2)^{1/2}}=n(n+1)D_n(x).$$

This not only establishes the result, but also shows that Turán's inequality

$$D_n(x)\geq 0,\qquad -1\leq x\leq 1,$$

with equality only for $x=\pm 1$) is a direct consequence of the fact that

$$|P_n(x)|<1,\qquad -1<x<1.$$

Making Nothing from Something

6520 [1986, 403]. *Proposed by Robert B. Israel, University of British Columbia.*

Show there exists a rational function f such that, for every holomorphic function ; in the unit disk D, g or $g-f$ has a zero in D. (See Problem 6437 [1983, 485; .985, 365].)

Solution by the proposer. By Runge's theorem, for every positive integer n there is a rational function f_n such that

$$|f_n(z)-z|<\frac{1}{n}\quad\text{for }\frac{1}{2}\leq|z|\leq 1 \tag{1}$$

$$\left|f_n(z)-\frac{1}{nz}\right|<\frac{1}{n}\quad\text{for }|z|=\frac{1}{4}. \tag{2}$$

Note that by the Argument Principle, f_n has a zero in $|z|<1/2$ and a pole in $|z|<1/4$. I claim that for sufficiently large n, f_n satisfies the requirement of the problem. Suppose this were false. Then there would be a sequence of functions g_n holomorphic in D, such that g_n and g_n-f_n have no zeros in D.

By Rouché's Theorem, for large n there must be z_n with $|z_n|=3/4$ and $|g_n(z_n)|>1/4$, since otherwise f_n-g_n and f_n would have the same number of zeros in $|z_n|<3/4$ (of course they have the same number of poles). Similarly, for large n there is w_n with $|w_n|=1/4$ and $|g_n(w_n)|<5/n$, since otherwise $1/f_n$ and

$1/f_n - 1/g_n$ would have the same number of zeros in $|z| < 1/4$. By the Maximum Modulus Theorem (applied to $1/g_n$) there must be p_n with $|p_n| = 3/4$ and $|g_n(p_n)| < 5/n$.

On the annulus $A = \{z: 1/2 < |z| < 1\}$ the holomorphic functions g_n/f_n do not take the values 0 and 1, and so by the Montel-Caratheodory Theorem some subsequence g_{n_j}/f_{n_j} converges uniformly on compact subsets of A, either to ∞ or to a holomorphic function on A. Since $g_{n_j}(p_{n_j})/f_{n_j}(p_{n_j}) \to 0$ it converges to a holomorphic function, and in fact (by Hurwitz's Theorem) to 0. But $g_{n_j}(z_{n_j})/f_{n_j}(z_{n_j})$ does not converge to 0, which is a contradiction.

Editorial Comment. This result may be stated as follows. The three functions f, 0, and ∞ form an "unavoidable family" of meromorphic functions in the interior of the unit disc. For the above terminology, and some related results (but in the whole plane $\mathbb{C}$ rather than the disc) see Lee A. Rubel and Chung-Chun Yan, Interpolation and unavoidable families of meromorphic functions, *Michigan Math. J.*, 20(1973) 289–296.

REVIEWS

EDITED BY JOSEPH KONHAUSER
Department of Mathematics, Macalester College, St. Paul, MN 55105

Studies in Mathematical Economics. MAA Studies in Mathematics, volume 25. Stanley Reiter, editor. The Mathematical Association of America, 1986, xiii + 441.

ROBERT A. BECKER
Department of Economics, Indiana University, Bloomington, IN 47405

1. Introduction. The problem of allocating scarce resources to achieve competing and insatiable ends lies at the core of economic reasoning. The analysis of the question "what's best" by mathematical means is a focal point of economic theory. Economists use a variety of mathematical techniques to approach this fundamental problem. Shifts in techniques come about in order to expand the theory as well to take advantage of the discovery of new mathematical methods. Calculus, set theory, topology, real and convex analysis, as well as probability theory play important roles in the mathematical expression of economic theory.

The classical Lagrange multiplier technique for examining local optima dominated the early development of the theory of consumer and producer behavior. This work dates to the 1930's and 1940's. In that period, Hicks and Samuelson advanced the frontier of the Walrasian general equilibrium framework by exploiting the calculus to explore both the stability of equilibrium and to perform the sensitivity analysis known as comparative statics. These two aspects of equilibrium theory form the basis for the predictive components of the Walras equilibrium theory.

Convex analysis methods first appeared in the seminal contributions of von Neumann to game theory in 1928 and the existence of an equilibrium configuration in a model of an expanding economy in 1938. Von Neumann's Minimax Theorem originates in those papers; it is one of the significant contributions of economic theory to mathematics.

During the Second World War, the mathematical focus shifted from the calculus to game theory and linear programming with the efforts of von Neumann, Morgenstern, Koopmans, and Dantzig (the independent contribution of Kantorovich to linear programming was not recognized until much later). Set theoretic and topological tools dominated the writings of mathematical economists in the 1950's and led to the fundamental properties of the Walrasian model of perfect competition: the existence of equilibrium and the welfare properties of equilibrium according to the Pareto criterion. Nonlinear and dynamic programming were also stimulated in part by the needs of economic theory in this period. Differential methods could still be found in the study of the uniqueness, stability, and comparative statics properties of equilibrium. Sufficient conditions for the uniqueness of a Walras equilibrium were known to be special.

Differential topology techniques were introduced into economics by Debreu in 1970. He showed how differentiability of demand could be used to show "most" (i.e, the regular) economies have only a finite number of equilibria. This new use of the calculus went beyond the methods of the 1930's and 1940's. Debreu's research in the 1950's influenced the shift away from the calculus as an important tool in favor of convex analysis. Ironically, Debreu was also the major influence in returning the calculus to favor in the 1970's.

Economic theory has taken an increasingly mathematical orientation since the Second World War. The variety of nonlinear problems found in economics has made mathematically inclined economists both end users of and suppliers to a variety of mathematical fields. It is now commonplace in the top economics journals to see papers employing differential topology, functional analysis (both linear and nonlinear theory), or stochastic processes. This raises the problem of how to bridge the communications gap between the two professions in an era of increasing specialization.

The volume *Studies in Mathematical Economics*, edited by Stanley Reiter, is an attempt to provide a series of papers by distinguished economic theorists which illustrate the scope and power of mathematical methods in economics. The papers range over game theory, the existence, uniqueness, and computation of Walras equilibrium, the nature and meaning of a decentralized resource allocation process, and the notion of "satisficing" rather than optimizing behavior. The mathematical methods include a sophisticated "advanced calculus" survey of constrained optimization problems defined on a finite dimensional vector space, the implicit function theorem, Sard's Theorem, the development of numerical algorithms for the computation of a fixed point of a nonlinear map (both a combinatorial approach keyed to Sperner's Lemma and a path following approach tied to Sard's Theorem are developed in separate chapters), the strong law of large numbers, and the theory of Markov chains with a countable state space. For the remainder of this essay, I will concentrate on the use of differential topology in economics and on issues in game theory in order to illustrate the variety of models, techniques, and results forming the subject of the Reiter volume.

2. Differential topology and the calculus. A demand function is defined as a mapping from an n-vector of prices into an n-vector of commodities. It represents the optimization choice of a consumer subject to a budget constraint. The laws of change stating how consumer demand for various goods responds to a change in market conditions may be found by calculating the derivative of the demand function with respect to the price vector.

The calculus approach to choice problems in economics is represented by Carl Simon's paper in this volume. The role of the implicit function theorem and the presentation of criteria for determining the global optimum are presented in detail. The range of methods and examples found in this chapter could be used to show how economic theory passes from postulates on behavior to model formulation and analysis. It would be an excellent source of economics examples of the calculus in action for students in multivariate and advanced calculus courses.

The differential calculus has a deeper use in models of the workings of the economy as a whole. A central problem in the study of the Walras general equilibrium model is to demonstrate the existence of a price system which equates supply and demand in all markets simultaneously. The simplest realization of this model is an exchange economy where the demand functions are derived by adding the demand functions of the individual consumers at different prices and defining the supply function to be the constant function equal to the total amount of each good initially available in the economy. The existence of a Walras equilibrium amounts to proving that a system of nonlinear equations in the unknown price system has at least one nonnegative solution. The existence task usually relies on a fixed point argument.

The next problem is to determine the number of possible solutions. Differential topology plays a decisive role in settling this question. An economy is defined for each vector of initial resources; the space of economies may be identified with the commodity space. An economy is *regular* if the demand function is differentiable and its Jacobian has maximal rank at every equilibrium price.

Let f be a smooth mapping defined on an open subset of $\mathbb{R}^m$ and taking values in $\mathbb{R}^n$. A point in the domain is a *critical point* if the Jacobian of f, evaluated at that point, has rank less than n. A point is *regular* if it is not a critical point. Sard's Theorem states that the image of the set of critical points of f has Lebesgue measure zero.

Debreu's fundamental result is based on Sard's Theorem and states that the nonregular economies form a closed set of Lebesgue measure zero in the space of economies. Moreover, every regular economy has an odd number of equilibria. The proof of the latter fact is based on an Index Theorem given in Mas–Collel's chapter, which also applies to some production economies.

The Index Theorem is closely related to the degree of a mapping. The index of a Walras equilibrium is the sign of the determinant of a matrix based on the Jacobian of the demand function at the given equilibrium; a normalization imposed by economic theory implies that the Jacobian of demand is a singular matrix, hence it cannot be directly used in the index calculation. The Index Theorem states that for every regular economy, the sum of the indices over all equilibria equals $(-1)^{n-1}$, where n is the number of commodities. The Index Theorem has two corollaries for regular economies: first, there is an odd number of Walras equilibria; in particular, a Walras equilibrium must exist. Second, if the index has a uniform sign across all equilibria, then there is only one equilibrium price system.

The Index Theorem and its corollaries appear to offer the prospect of deriving comparative statics results for the Walrasian economy. The implicit function theorem could be used for regular economies to compute the change in equilibrium prices in response to a parameter shift in the demand function for the economy. A series of results initiated by the work of Sonnenschein suggests that economic theory imposes very few restrictions on the form of the demand function in an exchange economy. The derivation of comparative statics results without further assumptions seems doubtful. One trend begun by Hildenbrand is to make stronger assumptions

on the distribution of preferences or income in order to obtain more characteristics of the economy-wide demand function.

One of the most important influences of the regular economies approach has been to focus researchers' attention on generic results. For example, the nonregular exchange economies are dismissed as unimportant examples. Regularity arguments can be found in the study of the determinateness of equilibrium in models motivated by fundamental problems in macroeconomic theory. Some other examples of the impact of differentiable methods can be found in Hurwicz's essay on decentralization. His and Reiter's papers are also good sources of other applications of general topological methods in economics.

3. Uncertainty and Information. Uncertainty is an important aspect of economic life. The nature and amount of information about the states of nature and the economy plays a fundamental role in understanding resource allocation processes. Information and uncertainty are dual concepts; to describe the information available to an agent is to describe the degree of uncertainty that the agent faces in the economic sphere. Information can enter an economic model in a variety of ways. Game theory provides one example.

The chapter by Roger Myerson on game theory attempts to illustrate the ways in which mathematical models can shed light on the possible equilibrium positions achievable under the hypothesis of rational agents pursuing their own ends within a world of like-minded individuals. The formulation of a game depends on a specification of the information available to each player as well as appropriate probabilities to be used in calculating expected utility payoffs. The expected utility theory forms the decision theoretic basis of the player's choice problems. The most important characteristic of the expected utility framework is that a player's payoff function is "linear in the probabilities" assigned to the various possible states of the world. One of the most interesting developments in economic theory in the 1980's is the exploration of non-additive expected utility theories by Machina, Chew, and Yaari. The interest in an alternative to the traditional expected utility theory arises from a controversy over its ability to predict behavior. There are several paradoxes which arise in the additive expected utility model which have been traced to the Independence Axiom. This postulate imposes the structure that yields the "additive" form of the utility function. The development of the implications of non-additive expected utility for game theory appears to be an area for promising research.

Given a finite set of *players* $N = \{1, 2, \ldots, n\}$ a *noncooperative game* is a triple $(N, u_n, S(n))$ where $S(n)$ is the set of strategies that are open to player n and u_n is a real-valued function on the cartesian product of the $S(n)$ that defines the payoff functions for player n as a function of the strategies selected by all the players. A necessary condition for any equilibrium strategy or outcome of the game is that a player's strategy has the property that he cannot increase his payoff by departing from his strategy as long as the other players maintain their strategies. A *Nash equilibrium* is a strategy for each player which is simultaneously chosen and self-enforcing in this sense. The fundamental result in game theory is the Nash

Theorem stating conditions for a noncooperative game to have an equilibrium solution. The existence theory for the Nash equilibrium concept applies in the additive expected utility framework; it should be applicable to the non-additive expected utility model as well.

Nash equilibria are abundant in many models; a substantial part of Myerson's paper is devoted to surveying the alternative ways of arriving at one equilibrium outcome. For example, information plays a decisive role in all of the noncooperative selection theories. This often changes the focus of the game form to an *extensive* or *multistage* form in contrast to the strategic form of the game postulated in the original Nash theory. Alternative equilibrium concepts depend on the specification of what each agent has in an information set and believes at each stage. For example, one could ask that the equilibrium outcome have the property that a player's strategies maximize the agent's conditionally expected payoff given his beliefs about the other players' actions and random events in any information set. A consistency rule also governs the players' beliefs at the various information states. The result of imposing these requirements is called a *sequential equilibrium*; it does give the researchers a nonempty subset of the Nash equilibrium set. Information holds the key to this advance. The additive expected utility hypothesis is also a critical component in the development of this theory. The impact of the non-additive theories on multistage games remains to be seen.

5. Conclusion. Stanley Reiter has brought a sample of the ways in which economists use mathematics together in one volume. The papers are all devoted to problems at the heart of economics. The subjects treated are not exhaustive, but this volume does serve as an introduction to mathematical reasoning in economics. For the teacher of mathematics, this volume opens a door to using economic theory as an alternative to the physical and biological applications of nontrivial mathematical analysis.

TELEGRAPHIC REVIEWS

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T:	Textbook	P:	Professional Reading	1–4:	Semesters
C:	Computer Software	L:	Undergraduate Library	**:	Special Emphasis
S:	Supplementary Reading	13:	Grade Level	??:	Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

General, S, P, L**.** *Riddles of the Sphinx And Other Mathematical Puzzle Tales.* Martin Gardner. New Math. Lib., V. 32. MAA, 1987, x + 164 pp, $14.50 (P). [ISBN: 0-88385-632-8] Thirty-six mathematical riddles embedded in imaginative stories first published in *Isaac Asimov's Science Fiction Magazine.* Riddles are followed by four sections of "answers" that unwrap successive layers of the puzzles, only to reveal hidden conundrums. Superb for browsing—one can plunge in anywhere and be instantly enthralled. Excellent supplement for mathematics clubs. LAS

General, P. *Fourteen Papers Translated from the Russian.* A. Ya. Aĭzenshtat. Transl: Ben Silver. AMS Transl. Ser. 2, V. 134. AMS, 1987, v + 153 pp, $49. [ISBN: 0-8218-3110-0]

Mathematics Appreciation, S(13-16), P, L**.** *Mathematics: Queen and Servant of Science.* E.T. Bell. MAA Spectrum. MAA, 1987, xxiv + 437 pp, $18.50 (P). [ISBN: 0-88385-447-3] First volume in MAA's new "Spectrum" series of popular mathematics books. This paperback is a reprint (with a Foreward by Martin Gardner) of Bell's 1951 classic, itself a revised composite of two earlier works on pure mathematics (the "queen") and applied mathematics (the "handmaiden") first published in 1931 and 1937. A master expositor, Bell cajoles the novice on a meandering tour of traditional mathematics, revealing beauty, application, and wit on every page. Should be required reading for every mathematics student. LAS

Mathematics Appreciation, T(13: 1). *Mathematics as a Second Language, Fourth Edition.* Joseph Newmark, Frances Lake. Addison-Wesley, 1987, xxi + 832 pp, $31.95. [ISBN: 0-201-05885-5] Well-written text introducing basic concepts. Excellent applications, examples, graphs and newspaper clippings from business, sociology, education, medicine. Historical notes and calculator exercises throughout. Topics include sets, logic, number systems, geometry, matrices, probability, statistics, and an introduction to BASIC. Good problems, chapter objectives and mastery tests. (*First Edition,* TR, May 1974; *Second Edition,* TR, November 1977; *Third Edition,* TR, December 1982.) MS

Logic, T(15-18: 1, 2), S, P. *Logics of Time and Computation.* Robert Goldblatt. CSLI Lect. Notes, No. 7. CSLI (Stanford U., Ventura Hall, Stanford, CA 94305), 1987, ix + 131 pp, $11.95 (P); $24.95. [ISBN: 0-937073-12-1; 0-937073-11-3] Discusses modal logic, emphasizing temporal and dynamic logics. Part one is on the basic theory of normal modal and temporal propositional logics. Part two is an application of this theory to temporal and computational logics, while part three is devoted to first-order dynamic logic. Exercises throughout. Bibliography. Index. RJA

Combinatorics, T(16-18: 1), S, P, L*. *Simulated Annealing: Theory and Applications.* P.J.M. van Laarhoven, E.H.L. Aarts. Math. & Its Applic. D Reidel (US Distr: Kluwer Academic), 1987, xi + 186 pp, $49. [ISBN: 90-277-2513-6] Simulated annealing is an efficient, iterative algorithm for minimizing the cost function in an NP-complete combinatorial problem (e.g., travelling salesman problem) by capitalizing on a deep and close analogy with annealing, the slow cooling of a solid until it reaches its low energy ground state. This monograph discusses the theoretical basis of the algorithm (asymptotic convergence), implementation performance, and a few

of the many dozens of applications. LAS

Number Theory, T(17-18: 1), S, P, L. *Elliptic Functions, Second Edition.* Serge Lang. Grad. Texts in Math., V. 112. Springer-Verlag, 1987, xi + 326 pp, $38. [ISBN: 0-387-96508-4] Elliptic functions live in the overlap between complex analysis and algebraic number theory. Lang treats general theory "from scratch," complex multiplication, nonintegral invariants, and the Kronecker limit formula. Slightly revised from the 1973 edition (TR, June-July 1974). BC

Number Theory, P. *Lecture Notes in Mathematics-1254: Explicit Constructions of Automorphic L-Functions.* Stephen Gelbart, Ilya Piatetski-Shapiro, Stephen Rallis. Springer-Verlag, 1987, vi + 152 pp, $18 (P). [ISBN: 0-387-17848-1] Two papers on L-functions attached to classical groups. The first applies a formal identity to the simple classical groups. The second generalizes the Rankin-Selberg method to certain non-simple groups of the form $G \times GL(n)$. BC

Number Theory, P. *Lecture Notes in Mathematics-1253: An Approach to the Selberg Trace Formula via the Selberg Zeta-Function.* Jürgen Fischer. Springer-Verlag, 1987, 184 pp, $18 (P). [ISBN: 0-387-15208-3] The Selberg zeta function is for Fuchsian groups what the Riemann zeta function is for the integers. The trace formula is a variant of Poisson summation, with miraculous implications for the zeta function (e.g., a "Riemann Hypothesis"). Fischer uses a "resolvent method" to clarify the miracle. BC

Group Theory, S(18), P. *Representations of Algebraic Groups.* Jens Carsten Jantzen. Pure & Appl. Math., V. 131. Academic Pr, 1987, xiii + 443 pp, $59.50. [ISBN: 0-12-380245-8] Part I, which requires knowledge of varieties and algebraic groups, develops the representation theory of algebraic group schemes. Part II covers representations of reductive groups, the prerequisite being the basics of reductive algebraic groups over an algebraically closed field. GG

Group Theory, T(18: 1). *Transformation Groups.* Tammo tom Dieck. Stud. In Math., V. 8. Walter de Gruyter, 1987, x + 312 pp, DM 128. [ISBN: 0-89925-029-7] An introduction to the theory of compact transformation groups, concentrating on the algebraic topology of Lie transformation groups. The reader is expected to have some understanding of basic algebraic topology. References to the literature provide background material. Many interesting exercises are included. RH

Group Theory, P. *Lecture Notes in Mathematics-1243: Non-Commutative Harmonic Analysis and Lie Groups.* Ed: J. Carmona, P. Delorme, M. Vergne. Springer-Verlag, 1987, 309 pp, $31.50 (P). [ISBN: 0-387-17701-9] Thirteen papers from the sixth colloquium on harmonic analysis and Lie groups, held at Marseille-Luminy in June 1985. GG

Algebra, T(18: 1), S, P. *Complex Semisimple Lie Algebras.* Jean-Pierre Serre. Springer-Verlag, 1987, ix + 74 pp, $19.80 (P). [ISBN: 0-387-96569-6] Based on lecture notes from 1965, there is a brief survey of the general theory of Lie algebras followed by a detailed discussion of complex semisimple algebras. Chapters on SL_2, root-systems, and structure theorems. Concludes with a chapter on compact groups. Bibliography, index. JS

Calculus, T(13), S, L. *Applied Calculus: An Intuitive Approach for Management, Life, and Social Sciences.* Richard L. Faber, Marvin I. Freedman, James L. Kaplan. West, 1986, xiii + 607 pp, $34.95. [ISBN: 0-314-85235-2] A text written for students preparing for careers in business, economics, or in the life or social sciences. The book is replete with examples from all these fields. The title is very apt, since the authors illustrate almost every concept with several intuitive examples designed to help students understand the "why" of calculus. Contains sections on mathematical modeling, numerical integration, price elasticity, least squares, Lagrange multipliers, power series solution of differential equations, Newton's method and a chapter on probability. Trigonometric functions are introduced following differentiation and integration in several variables; exponential and logarithmic functions are introduced much earlier. SM

Calculus, P. *Zahlen Kontinuum: Eine Einführung in die Infinitesimalmathematik.* Detlef Laugwitz. Lehrbucher und Mono. zur Didaktik der Math., B. 5. Bibliographisches Institut, 1986, 269 pp, 38 DM (P). [ISBN: 3-411-03128-X] An introduction to nonstandard analysis, based on the adjunction of an ideal element to the reals and intended for teachers. Many historical and philosophical remarks. JD-B

Differential Equations, P. *Hydrodynamic Behavior and Interacting Particle Systems.* Ed: George Papanicolaou. Inst. for Math. & Its Applic., V. 9. Springer-Verlag, 1987, xi + 210 pp, $19.80. [ISBN: 0-387-96584-X] Proceedings of a March 1986 workshop conducted as part of the 1985-86 IMA program on "Stochastic Differential Equations and Their Applications." Contains fifteen papers presenting experimental results on suspensions and a discussion of analytical methods used to study these systems. AM

Partial Differential Equations. *Partial Differential Equations.* J. Wloka. Transl: C.B. & M.J. Thomas. Cambridge U Pr, 1987, xi + 518 pp, $79.50; $29.95 (P). [ISBN: 0-521-25914-2; 0-521-27759-0] Discusses theoretical and numerical aspects of solving boundary value problems. Assumes familiarity with basic functional analysis and begins with an introduction to distributions and Sobolev spaces. With this foundation, the book presents different methods for solving elliptic, parabolic, and hyper-

bolic equations, as well as difference methods for numerical solution of partial differential equations. AM

Partial Differential Equations, P. *Problèmes de Dirichlet Variationnels non Linéaires.* Jean Mawhin. Pr U Montreal, 1987, 168 pp, $22 (P). [ISBN: 2-7606-0799-2]

Partial Differential Equations, P. *Nonlinear Parabolic Equations: Qualitative Properties of Solutions.* Ed: L. Boccardo, A. Tesei. Res. Notes in Math. Ser., V. 149. Longman Scientific & Technical (US Dist: Wiley), 1987, 232 pp, $49.95 (P). [ISBN: 0-470-20379-X] Proceedings of a conference held at the Second Rome University from April 1-5, 1985. The book contains 31 papers which discuss the following topics: existence, uniqueness and regularity of solutions for parabolic equations and systems; special classes of solutions; free boundary problems; degenerate diffusion; asymptotical behaviour of solutions; and models from biology, physics, and engineering. AM

Partial Differential Equations, P. *Lecture Notes in Mathematics-1256: Pseudo-Differential Operators.* Ed: H.O. Cordes, B. Gramsch, H. Widom. Springer-Verlag, 1987, x + 479 pp, $41.70 (P). [ISBN: 0-387-17856-2] Twenty-two research papers on pseudodifferential operators, a spin-off from partial differential equations. Includes a lengthy contribution by Hörmander on non-linear hyperbolic equations. BC

Partial Differential Equations, P. *Oakland Conference on Partial Differential Equations and Applied Mathematics.* Ed: Louis R. Bragg, John W. Dettman. Res. Note in Math. Ser., V. 154. Longman Scientific & Technical (US Distr: Wiley), 1987, 117 pp, $38.95 (P). [ISBN: 0-470-20777-9] Five papers from a conference at Oakland University, Michigan, focusing on transmutations and other methods in partial differential equations. Topics include inverse scattering theory, ocean acoustics, and potential theory. BC

Partial Differential Equations, P. *Lecture Notes in Mathematics-1260: Nonlinear Evolution Operators and Semigroups.* Nicolae H. Pavel. Springer-Verlag, 1987, vi + 285 pp, $24.30 (P). [ISBN: 0-387-17974-7] An exposition of fundamental results and recent research on nonlinear evolutionary operators and semigroups, with applications of the theory to several standard types of differential equations arising in physics. The techniques allow partial differential equations to be treated as ordinary differential equations in infinite dimensional Banach spaces, enabling a unified approach. RB

Partial Differential Equations, P. *Lecture Notes in Mathematics-1236: Stochastic Partial Differential Equations and Applications.* Ed: G. Da Prato, L. Tubaro. Springer-Verlag, 1987, 257 pp, $23.60 (P). [ISBN: 0-387-17211-4] Proceedings of an October 1985 conference in Trento, Italy. LAS

Numerical Analysis, T(16-17: 1, 2), S*, L. *Numerical Methods in Engineering and Applied Science: Numbers are Fun.* B. Irons, N.G. Shrive. Math. & Its Applic. Halsted Pr, 1987, 248 pp, $32.95. [ISBN: 0-470-20803-1] A delightful, witty, back-of-the-envelope survey of how engineers think about mathematical methods intended to fill "regrettable gaps" in students' education. Covers integration, interpolation, splines, differential equations, matrices, eigenvalues—not systematically but in the manner of a personalized idiosyncratic tour. Excellent source of intuition and synthetic ideas. LAS

Numerical Analysis, P. *Defect Minimization in Operator Equations: Theory and Applications.* R. Reemtsen. Res. Notes in Math. Ser., V. 163. Longman Scientific & Technical (US Distr: Wiley), 1987, 106 pp, $34.95 (P). [ISBN: 0-470-20877-5] Defect minimization is a numerical technique for approximating solutions to operator equations $Ta = r$ by minimizing $Ta - r$ (the defect) in finite-dimensional subspaces for a. This book develops the general theory (especially convergence), and applies it to ordinary differential equations and the heat equation. Begins with a wonderful quote from *A Hundred Years of Solitude.* BC

Numerical Analysis, T(16-17: 1), S, P, L. *Sensitivity Analysis in Linear Systems.* Assem Deif. Springer-Verlag, 1986, xii + 224 pp, $56. [ISBN: 0-387-16312-3] Five self-contained chapters (with thorough examples and exercises) on perturbation techniques and sensitivity analysis: perturbation of linear equations, interval analysis, iterative systems, least-squares, and sensitivity in linear programming. BC

Functional Analysis, P. *Lecture Notes in Mathematics-1235: Séminaire de Théorie du Potentiel, Paris, No. 8.* Ed: F. Hirsch, G. Mokobodzki. Springer-Verlag, 1987, 209 pp, $19.40 (P). [ISBN: 0-387-17210-6]

Functional Analysis, T?(18: 2), P. *Barrelled Locally Convex Spaces.* Pedro Pérez Carreras, José Bonet. Math. Stud., V. 131. Elsevier Science, 1987, xv + 512 pp, $80 (P). [ISBN: 0-444-70129-X] This systematic treatment of barrelled spaces begins with background material from various branches of analysis and then focuses on Baire spaces. In addition to the abstract theory of barrelled spaces, the volume discusses local completeness, the inherited Mackey topology, bornological and ultrabornological spaces. Clearly written with extensive bibliography but no exercises. MU

Functional Analysis, S(17-18), P. *Inner Product Structures: Theory and Applications.* Vasile Ion Istrătescu. Math. & Its Applic. D Reidel (US Distr: Kluwer Academic), 1987, xv + 895 pp, $149. [ISBN: 90-277-2182-3] A thorough, self-contained de-

velopment of inner product structures, which underlie many linear spaces. Applications to probability and physics interspersed. An extensive bibliography and a collection of open problems. BC

Analysis, P. *Fourteen Papers Translated from the Russian.* L.A. Aksent'ev, *et al.* Transl: Ben Silver. AMS Transl. Ser. 2, V. 136. AMS, 1987, v + 154 pp, $47. [ISBN: 0-8218-3112-7] Thirteen papers on analysis and one on social choice theory. LAS

Analysis, P. *Lecture Notes in Mathematics-1262: Zahlentheoretische Analysis II.* E. Hlawka. Springer-Verlag, 1987, 158 pp, $16.30 (P). [ISBN: 0-387-18015-X]

Algebraic Geometry, P. *Lecture Notes in Mathematics-1266: Space Curves.* Ed: F. Ghione, C. Peskine, E. Sernesi. Springer-Veralg, 1987, 272 pp, $24.30 (P). [ISBN: 0-387-18020-6] Proceedings of a conference in Rome, June 1985 on "Curves in Projective Space." Twelve papers: "good and bad families of projective curves;" postulation of projective space curves; classical problems in enumerative geometry. RB

Differential Geometry, P. *Lecture Notes in Mathematics-1263: Differential Geometry.* Ed: V.L. Hansen. Springer-Verlag, 1987, xi + 288 pp, $24.30 (P). [ISBN: 0-387-18012-5] Proceedings of the Nordic Summer School in mathematics held at the Technical University of Denmark in Lyngby. Contains the manuscripts for the main lectures given at the school. These lectures discussed gauge theory and moduli; twistor methods; and global differential geometry. AM

Differential Geometry, P. *Monopoles, Minimal Surfaces and Algebraic Curves.* Nigel Hitchin. Pr U Montreal, 1987, 94 pp, $18 (P). [ISBN: 2-7606-0801-8] Monograph studies a collection of non-linear differential equations that derive from the self-dual Yang-Mills equations. The techniques of symplectic geometry and moment maps are used to study the structure of the moduli space of monopoles—a special class of solutions to the Yang-Mills equations. AM

Algebraic Topology, P. *Homotopy Theory and Related Topics.* Ed: H. Toda. Adv. Stud. in Pure Math., V. 9. Elsevier Science, 1987, xii + 344 pp, Dfl. 250.00. [ISBN: 0-444-70201-6] Proceedings of a Japanese symposium (Kyoto, December 1984). Nineteen research papers and four expository lectures on simple homotopy theory and G-actions, classifying spaces and characteristic classes, topology of manifolds, homotopy problems (unstable and stable cases). RB

Dynamical Systems, P. *Dynamical Systems and Bifurcation Theory.* M.I. Camacho, M.J. Pacifico, F. Takens. Res. Notes in Math. Ser., V. 160. Longman Scientific & Technical (US Distr: Wiley), 1987, 421 pp, $62.95 (P). [ISBN: 0-470-20843-0] Proceedings of a conference at Rio de Janeiro, August 1985. Thirteen papers, primarily on structural stability of vector fields and bifurcation theory, including Hopf bifurcation, transition from periodic to strange attractors and homoclinic tangencies. Other topics: polynomial foliations, complex vector fields and foliations, dynamical aspects of lines of curvature; geodesic flows. RB

Dynamical Systems, S(18). *Lecture Notes in Control and Information Sciences-77: Detection of Abrupt Changes in Signals and Dynamical Systems.* Ed: M. Basseville, A. Benveniste. Springer-Verlag, 1986, x + 373 pp, $28.50 (P). [ISBN: 0-387-16043-4] The detection of abrupt changes in dynamical systems has applications which range from the detection of sensor failure to the prediction of the catastrophic failure of mechanical and biological systems. Certain meteorological phenomena, tsunamis for example, can also be considered abrupt changes in dynamic systems. Hypothesis testing plays an important role in the detection of these changes. This book presents a series of twelve research papers which cover the topic from both a theoretical and applied perspective. SM

Control Theory, P. *Algebraic and Geometric Methods in Nonlinear Control Theory.* Ed: M. Fliess, M. Hazewinkel. Math. & Its Applic. D Reidel (US Distr: Kluwer Academic), 1986, xii + 642 pp, $99. [ISBN: 90-277-2286-2] The proceedings of a conference sponsored by the Centre National de la Recherche Scientifique and the Laboratorie des Signaux et Systems, held in Paris in June 1985. Contains 31 papers on state-of-the-art nonlinear control theory. Topics include controllability, observability, and realization; feedback synthesis and linearization; optimal control; discrete time systems; and applications. Survey papers and new results are presented. SM

Control Theory, P. *Lecture Notes in Control and Information Sciences-91: Stochastic Modelling and Filtering.* Ed: Alfredo Germani. Springer-Verlag, 1987, 218 pp, $35.60 (P). [ISBN: 0-387-17575-X] Proceedings of a December 1984 IFIP workshop held in Rome: fourteen papers on recent advances in theory and applications of stochastic modelling. LAS

Control Theory, P. *Modelling, Robustness and Sensitivity Reduction in Control Systems.* Ed: Ruth F. Curtain. NATO ASI Ser. F, V. 34. Springer-Verlag, 1987, ix + 492 pp, $89. [ISBN: 0-387-17845-7] Papers on the theory and practice of modelling of control systems. Mathematical foundations of H^∞-control theory, modelling from measurement data, approximation, model reduction, robustness and sensitivity reduction. RM

Systems Theory, T(18: 1), P. *Stability of Adaptive Systems: Passivity and Averaging Analysis.*

B.D.O. Anderson, *et al.* Ser. in Signal Proc., Optimiz. & Control., V. 8. MIT Pr, 1986, xviii + 326 pp, $35. [ISBN: 0-262-01090-9] This book, with eight authors, consists of six chapters on the stability of adaptive systems, systems which depend on a set of time varying parameters. The ultimate goal of this field is to develop algorithms which remain stable while the system evolves into the future. This book concerns itself more with methodology and basic concepts rather than with the specific details of any algorithm. The final chapter includes a detailed example. SM

Probability, T*(15-16: 1, 2), S, L. *Probability and Stochastic Processes.* Frederick Solomon. Prentice-Hall, 1987, xv + 426 pp. [ISBN: 0-13-711961-5] An introduction to probability at a level that takes the axiom of countable additivity as optional. Includes discussions of correlation coefficients, the central limit theorem, continuous-time birth and death processes, and discrete-time Markov chains. FLW

Probability, S*, P, L.** *Paradoxes in Probability Theory and Mathematical Statistics.* Gábor J. Székely. Math. & Its Applic. D Reidel (US Distr: Kluwer Academic), 1986, xii + 250 pp, $59. [ISBN: 90-277-1899-7] A fascinating exposition of paradoxical results in probability, statistics, random processes, and foundations of probability, complete with historical detail on the origin and resolution of each paradox. Topics range from old dilemmas about dice that led to the creation of probability theory to trapdoor functions in cryptography. Historical context shows how often paradoxes led to new insights in the mathematics of chance. LAS

Stochastic Processes, P. *Random Media.* Ed: George Papanicolaou. IMA, V. 7. Springer-Verlag, 1987, xi + 321 pp, $34. [ISBN: 0-387-96524-6] Twenty-one papers from an IMA workshop, mostly on wave equations and random Schrödinger operators. One-dimensional problems are emphasized (in part because they can be solved). BC

Stochastic Processes, S(17). *Statistical Estimation for Stochastic Processes.* K. Nanthi, M.T. Wasan. Papers in Pure & Appl. Math., No. 78. Queen's U, 1987, vi + 245 pp, (P). Provides mathematical background, including estimation and test of hypothesis in discrete time Markov chains, maximum likelihood estimators, etc., for such problems as birth and death models, immigration and emmigration, diffusion, continuous time processes, and filtering. Statistical estimations for these problems aid in work where actual parameters are unknown, such as life table analysis (actuarial) and network theory (computer science). LB-E

Elementary Statistics, T*(13: 1, 2). *Exploring Statistics: A Modern Introduction to Data Analysis and Inference.* Larry J. Kitchens. West, 1987, xii + 595 pp, $30.56 [ISBN: 0-314-28498-2]; *Instructor's Manual with Solutions and Test Bank,* v + 333 pp, (P). [ISBN: 0-314-35115-9] Covers the usual topics, but in a more modern way than most introductory texts. Exploratory data analysis (EDA) techniques are introduced early and used throughout, robust and nonparametric procedures are integrated with standard methods, many examples and exercises use real data, and optional sections on using Minitab are included in each chapter. RSK

Elementary Statistics, S*(14-18), P*, L*. *A Handbook of Introductory Statistical Methods.* C. Philip Cox. Ser. in Prob. & Math. Stat. Wiley, 1987, xxi + 272 pp, $34.95. [ISBN: 0-471-81971-9] An "economical summary" of many statistical techniques for estimation and hypothesis testing. Elementary procedures, contingency tables, bivariate and multiple regression, multiple comparisons, correlations, and randomized block and Latin square designs. FLW

Statistics, S(15-18). *The Statistical Consultant in Action.* Ed: D.J. Hand, B.S. Everitt. Cambridge U Pr, 1987, xii + 189 pp, $34.50. [ISBN: 0-521-30717-1] Twelve articles, ranging from the entertaining to the technical, designed to give a realistic picture of statistical consulting. Shows the diversity of problems and challenges with which statisticians must deal. Includes a bibliography. RSK

Statistics, T(18: 1), P*. *Statistical Analysis with Missing Data.* Roderick J.A. Little, Donald B. Rubin. Prob. & Math. Stat. Wiley, 1987, xiv + 278 pp, $34.95. [ISBN: 0-471-80254-9] Comprehensive treatment for the applied statistician, divided into two parts. The shorter Part I reviews historical approaches to missing data problems in analysis of variance, survey sampling, and multivariate analysis. Part II presents a systematic approach using likelihood-based methods which take into account the mechanisms that lead to missing data. Includes applications to a wide variety of situations. RSK

Statistics, T(15-18: 1), S, P. *The General Linear Model: Data Analysis in the Social and Behavioral Sciences.* Raymond L. Horton. Robert E Krieger, 1986, xi + 274 pp, $29.95. [ISBN: 0-89874-906-9] A reprint of the 1978 edition. Develops the general linear model in matrix form, assumptions needed for inference, and use of the model for factorial, Latin square, repeated measures, covariance, and regression analysis designs. Written for the "applied researcher." FLW

Statistics, T(16-18: 1-3), S. *Statistical Models in Applied Science.* Karl V. Bury. Robert E Krieger, 1986, xvii + 625 pp, $49.50. [ISBN: 0-89874-747-3] A reprint of the 1975 edition (TR, October 1977). Statistical notions pertinent in engineering. Includes chapters on probabilistic design, component reliability, systems analysis, and strength of materials. FLW

Statistics, S(17-18), P. *Goodness-Of-Fit.* Ed: P. Révész, K. Sarkadi, P.K. Sen. Elsevier Science, 1987, 624 pp, $120. [ISBN: 0-444-70087-0] 36 papers presented at a colloquium held in Hungary in 1984 on the theory and application of goodness-of-fit tests. FLW

Statistics, T(15-16: 1), S. *Multivariate Data Analysis.* William W. Cooley, Paul R. Lohnes. Robert E Krieger, 1985, x + 364 pp, $30.50. [ISBN: 0-89874-781-3] A reprint of the 1971 edition. A data analytic and non-mathematical approach to factor analysis, canonical correlation, partial correlation, multivariate analysis of variance, and discriminant analysis. FLW

Statistics, S(17). *Mixture Models: Inference and Applications to Clustering.* Geoffrey J. McLachlan, Kaye E. Basford. Stat., V. 84. Dekker, 1988, xi + 253 pp, $69.75. [ISBN: 0-8247-7691-7] Theoretical and practical issues of clustering. Numerous examples of real data sets taken from agriculture, botany, education, medicine and zoology used to demonstrate applications of finite mixture models. Excellent source of references on work in area. Fortran listing of computer programs given. Good supplement for multivariate analysis texts. MS

Statistics, T(17-18: 1), S, P*, L. *Multiple Imputation for Nonresponse in Surveys.* Donald B. Rubin. Ser. in Prob. & Math. Stat. Wiley, 1987, xxix + 258 pp, $32.95. [ISBN: 0-471-08705-X] Gives a method for handling nonresponses in sample surveys by "replacing each missing value with two or more imputations, analyzing the resulting multiply-imputed data set using standard methods, and then combining the complete-data answers in simple ways." FLW

Elementary Computer Science, T(13-14). *Structured Programming Using True BASIC: An Introduction.* Wade Ellis, Jr., Ed Lodi. Harcourt Brace Jovanovich, 1988, xviii + 356 pp, $19 (P). [ISBN: 0-15-584076-2] Requires Macintosh, IBM (or compatible) PC's and True BASIC (a new BASIC incorporating structured programming technique not present in BASIC). Many sample programs. Problems at the end of each chapter. Appropriate for teaching structured programming on PC's without Pascal or other language. LB-E

Programming. *Complete BASIC for the Short Course.* James S. Quasney, John Maniotes. Boyd & Fraser, 1985, xi + 188 pp, $7.50 (P). [ISBN: 0-87835-151-5] Written to supplement a general textbook in an introductory data processing course with a BASIC programming component. Assumes no previous programming experience. Seven different dialects of the BASIC language are documented. Emphasizes structured programming techniques, minimizing use of GOTO statement. Includes many simple examples of applications. SM

Programming, T(13-18: 1), S. *The SCHEME Programming Language.* R. Kent Dybvig. Prentice-Hall, 1987, xi + 242 pp, $21.33 (P). [ISBN: 0-13-791864-X] Dialect of LISP supporting static scoping, first-class procedures, and continuations. Presupposes familiarity with Scheme or another programming language. Topics include an introduction for the novice Scheme programmer, creation and assignment of identifier bindings, program control operations, operations on object types, I/O operations, system operations, syntactic extension and structure definition. Ends with a chapter of longer, complete examples. Text uses the *Chez* Scheme implementation. References. Summary of Forms. Index. RJA

Programming, S(14-16), P*, L. *Elements of C.* Morton H. Lewin. Found. of Comput. Sci. Plenum Pr, 1986, xiv + 246 pp, $29.50. [ISBN: 0-306-42182-8] An exposition of the C programming language (as opposed to programming technique) which presents expressions (pointers, structures and all) as a whole before turning to program structure and complete programs. Expressly for the professional, experienced student, or hobbyist who already programs in assembly language and at least one high-level language. At last! RB

Programming, T(14-16: 1, 2), S. *A Structured Approach to Programming, Second Edition.* Joan K. Hughes, Glen C. Michtom, Jay I. Michtom. Prentice-Hall, 1987, x + 322 pp, $26.67. [ISBN: 0-13-854159-0] Topics include program design, planning and implementation, structured programming, stepwise refinement, structured walk-throughs and inspections, testing. Four language chapters, two of which are new on Pascal and on C. Also new material added on top-down design and testing. Chapter summaries, review questions and exercises. Glossary; bibliography; index. RJA

Programming, T(13-18: 1), S. *Advanced Turbo Prolog: Version 1.1.* Herbert Schildt. Osborne McGraw-Hill, 1987, xiii + 323 pp, $21.95 (P). [ISBN: 0-07-881285-2] Considers many typical artificial intelligence problems and their solutions. Chapters include an overview of artificial intelligence, search, expert systems, natural languages, vision and pattern recognition, robotics, machine learning, logic. Appendices on interfacing with other languages and on the Turbo Prolog Toolbox. Index. RJA

Programming, T(13-18: 1), S. *The T Programming Language: A Dialect of LISP.* Stephen Slade. Prentice-Hall, 1987, xiv + 425 pp, $21.33 (P). [ISBN: 0-13-881905-X] T is a version of the Scheme dialect of LISP. Assumes familiarity with programming in a language other than LISP. Suitable as a supplemental text for a course in artificial intelligence, data structures, computer systems, compiler design. In addition to regular LISP features, T includes procedural arguments, debugging and error checking, data abstraction, object-oriented programming, locales, op-

timizing compiler, uniform semantics for interpreted and compiled code. Chapter summaries. Answers to selected exercises. References. Index. RJA

Programming, T(13-18: 1), S. *The XLISP Primer.* Bonnie J. Fladung. Prentice-Hall, 1987, xix + 172 pp, $25.67 (P). [ISBN: 0-13-972084-7] A tutorial and reference manual for the XLISP interpreter. Provides an introduction to LISP, emphasizing functions compatible with Common LISP. Chapter exercises with answers. Summary of XLISP features; appendix of graphics and interface commands for the IBM PC and for the Macintosh. Index. RJA

Languages, T(13-18: 1), S*, P, L. *Productive Prolog Programming.* Peter Schnupp, Lawrence W. Bernhard. Prentice-Hall, 1987, xiv + 296 pp, $26 (P). [ISBN: 0-13-725110-6] Treats Prolog as a production language, as a tool for software engineers. Prolog is first viewed as a relational knowledge base, secondly as a procedural interpreter of its rules, thirdly as a pattern matcher in the form of the unification process. Subsequent to these three views, Prolog is investigated as a system of programming in logic. Application and implementation details, index. RJA

Computer Systems. *Lecture Notes in Computer Science-266: Advances in Petri Nets 1987.* Ed: Grzegorz Rozenberg. Springer-Verlag, 1987, vi + 451 pp, $34.60 (P). [ISBN: 0-387-18086-9] Papers from the seventh European workshop on applications and theory of Petri nets held in Oxford in June 1986. Concludes with a major comprehensive bibliography on Petri nets. LAS

Computer Graphics, P. *Techniques for Computer Graphics.* Ed: David F. Rogers, Rae A. Earnshaw. Springer-Verlag, 1987, viii + 512 pp, $59.50. [ISBN: 0-387-96492-4] Proceedings of the International Summer Institute on the State-of-the-Art in Computer Graphics held at the University of Stirling. Contains nineteen papers covering: workstations, graphics standards, image generation, computer-aided design, curves and surfaces, human-computer interface issues, electronic documents, integrated graphics and text, solid modeling, and VLSI. Many of the papers contain background discussions, so the book should be useful for non-specialists as well as for more experienced researchers. AM

Computer Graphics, P. *Advances in Computer Graphics: Hardware I.* Ed: W. Straßer. EurographicSeminars. Springer-Verlag, 1987, x + 145 pp, $44. [ISBN: 0-387-18222-5] Proceedings of a 1986 workshop. Contributions discuss workstation architectures, hardware support for geometric modelling, 3-D models, graphics for traffic and movement simulation, and ray tracing. RM

Theory of Computation, T(18: 1, 2), S, P. *Varieties of Formal Languages.* J.E. Pin. Transl: A. Howie. Plenum Pr, 1986, x + 138 pp, $37.50. [ISBN: 0-306-42294-8] Presents basic results on theory of finite automata and rational languages. The concept of variety of languages is used to formalize the relationship among finite automata, recognizable languages, finite semigroups. Chapter problems. Bibliographic notes. Index. RJA

Theory of Computation, P. *Lecture Notes in Computer Science-271: From Logic Design to Logic Programming.* Dominique Snyers, André Thayse. Springer-Verlag, 1987, iv + 125 pp, $15.40 (P). [ISBN: 0-387-18217-9] *P*-functions are described and related to the design of algorithms and to declarative programming. Topics include theorem proving, grammars, logics, semantics. Bibliography. RJA

Theory of Computation, T(18: 1), P. *The Complexity of Boolean Functions.* Ingo Wegener. Ser. in Computat. Sci. Wiley, 1987, xi + 457 pp, $44.95. [ISBN: 0-471-91555-6] Covers models for computing Boolean functions, parameters for measuring complexity, bounds on complexity, relationship of circuit models and Turing machines. Assumes basic knowledge of linear algebra, combinatorics, algorithm analysis, complexity theory. Exercises at end of each chapter. Extensive bibliography. KS

Theory of Computation, S(17). *Essays on Concepts, Formalisms, and Tools.* Ed: P.R.J. Asveld, A. Nijholt. CWI Tract, V. 42. Math Centrum, 1987, v + 278 pp, Dfl. 42.90 (P). [ISBN: 90-6196-326-5] This is a series of about a dozen papers addressing research issues in the field of theoretical computer science. These papers were presented at a symposium on this subject at Twente University, The Netherlands, in October 1987. The papers address such issues as formal languages, logic programming, parses, and formal models of computation. GMS

Theory of Computation, T(16-18: 1, 2), S, P, L. *Computability Theory, Semantics, and Logic Programming.* Melvin Fitting. Logic Guides, V. 13. Oxford U Pr, 1987, xi + 198 pp, $35. [ISBN: 0-19-503691-3] Early chapters investigate theoretical properties of the high-level language EFS. A model computer is subsequently introduced with a process for translating from EFS into its machine language. EFS is chosen to be a logic programming language, similar to Prolog. The model of a computer is based on register machines. Three major themes are developed: robustness, essential limitations, and program semantics. Exercises. Chapter background sections. Chapter references. Appendex on induction. Index. RJA

Artificial Intelligence, P. *Artificial Intelligence and Instruction: Applications and Methods.* Ed: Greg Kearsley. Addison-Wesley, 1987, xiv + 351 pp, $29.95. [ISBN: 0-201-11654-5] Individual chapters written by experts on different aspects of Intelligent Computer Assisted Instruction (ICAI). Begins with an overview of ICAI. Parts two and three concern ar-

tificial intelligence in education and in training. Part four is on building intelligent tutors, while the last part is on implementing ICAI systems. Chapter references. Index. RJA

Artificial Intelligence, T*(15-18: 1, 2), S, P, L. *Logical Foundations of Artificial Intelligence.* Michael R. Genesereth, Nils J. Nilsson. Morgan Kaufmann, 1987, xviii + 405 pp, $36.95. [ISBN: 0-934613-31-1] Text rests on two main assumptions: progress in a discipline requires development and use of appropriate mathematical apparatus to express ideas, and symbolic logic forms, one of the most important aspects of the mathematics for artificial intelligence. Topics include declarative knowledge, inference, resolution, nonmonotonic reasoning, induction, reasoning with uncertainty, knowledge and belief, planning, metaknowledge and metareasoning, state and change, intelligent-agent architecture. Chapter exercises, bibliographical and historical remarks. Answers to exercises. RJA

Applications, T(15: 1), L. *Introduction to Analytical Dynamics.* N.M.J. Woodhouse. Oxford U Pr, 1987, x + 169 pp, $39.95. [ISBN: 0-19-853198-2] A textbook for undergraduates. Slim, no-fat coordinate-based treatment of Newtonian mechanics. Requires background in linear algebra, multivariable calculus and some three-dimensional mechanics. Builds a case for Newtonian mechanics. Many exercises and examples from recent Oxford examination papers. JK

Applications (Electrical Engineering), T(14-16). *Fundamentals of Modern Digital Systems, Second Edition.* B.R. Bannister, D.G. Whitehead. Springer-Verlag, 1987, ix + 304 pp, $29.50. [ISBN: 0-387-91311-4] Good introduction to digital systems with excellent introductory chapters on combinational logic, numbering systems, and coding. Problems of varying levels with solutions to selected (not odd) problems. Fine illustrations well-spaced within text. Includes chapters on memory systems and programmable devices as well as more fundamental topics: semiconductors, logic design techniques, logic components and state tables. LB-E

Applications (Engineering), P, L. *Formulas, Facts and Constants for Students and Professionals in Engineering, Chemistry and Physics, Second, Revised and Enlarged Edition.* H.J. and K.H. Fischbeck. Springer-Verlag, 1987, xv + 260 pp, $23 (P). [ISBN: 0-387-11315-0]

Applications (Engineering), P. *Lecture Notes in Engineering-26: Shell and Spatial Structures: Computational Aspects.* Ed: G. De Roeck, *et al.* Springer-Verlag, 1987, vii + 487 pp, $60 (P). [ISBN: 0-387-17498-2] Proceedings of a July 1986 symposium held in Lenven, Belgium on new techniques for design of structures using engineering workstations. LAS

Applications (Physics). *Analysis and Mathematical Physics.* Hans Triebel. Math. & Its Applic. D Reidel, 1986, xxviii + 456 pp, $89. [ISBN: 90-277-2077-0] Lists without proof many major theorems from a vast area of mathematics, beginning with elementary algebra and trigonometry and progressing through differential and integral calculus of one and several variables, ordinary and partial differential equations, calculus of variations, measure and integration theory, complex function theory, Hilbert spaces, and much more. Apart from the definitions and statements of theorems, the text consists primarily of numerous disconnected remarks. MU

Applications (Physics), S(17-18), P. *Obstacle Problems in Mathematical Physics.* José-Francisco Rodrigues. Math. Stud., V. 134. Elsevier Science, 1987, xv + 352 pp, $71 (P). [ISBN: 0-444-70187-7] Presents a general account of the applicability of elliptic variational inequalities to obstacle type boundary problems. Beginning with some obstacle problems which can be reduced to variational inequalities, the book then presents some of the main aspects of elliptical inequalities from the abstract Hilbertian framework to the smoothness of the variational solution. Clearly written and laced with informal discussion. MU

Applications (Physics). *Nonlinear Diffusive Waves.* P.L. Sachdev. Cambridge U Pr, 1987, vii + 246 pp, $49.50. [ISBN: 0-521-26593-2] "This monograph deals with Bergers' equation and its generalizations. Such equations describe a wide variety of nonlinear diffusive phenomena, for instance, in nonlinear acoustics, laser physics, plasmas and atmospheric physics ...." Both analytical and numerical perspectives. Prerequisites: basic course in gas dynamics; ordinary and partial differential equations; parabolic partial differential equations helpful. RB

Reviewers

RJA: Richard J. Allen, St. Olaf; DFA: David F. Appleyard, Carleton; RB: Richard Brown, Carleton; LB-E: Lisa Brown-Ewing, Macalester; JNC: Judith N. Cederberg, St. Olaf; LC: Laura Chihara, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; RSF: Robert S. Fisk, St. Olaf; SG: Steven Galovich, Carleton; GG: George Gilbert, St. Olaf; BH: Bruce Hanson, St. Olaf; RH: Rhonda Hatcher, St. Olaf; RK: Roger Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; AM: Alan Magnuson, St. Olaf; SM: Steve McKelvey, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; MR: Matthew Richey, St. Olaf; AWR: A. Wayne Roberts, Macalester; GMS: G. Michael Schneider, Macalester; PS: Paul Schoening, St. Olaf; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MS: Myriam Steinback, Macalester; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; LW: Lynne Walling, St. Olaf; FLW: Frank L. Wolf, Carleton; MZ: Michael Zielinski, St. Olaf.

THE MATHEMATICAL ASSOCIATION OF AMERICA
1225 Connecticut Avenue, N.W.

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 95, Number 4 April 1988

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Dept. of Mathematics, Stanford University, Stanford, CA 94305 (address effective until May, 1988). Shorter articles about the teaching of mathematics should be sent to STAN WAGON and JOAN HUTCHINSON, Department of Mathematics, Smith College, Northampton, MA 01063. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to ANITA E. SOLOW, Dept. of Mathematics, Grinnell College, Grinnell, IA 50112; in geometry or topology to DENNIS DETURCK, Dept. of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to RICHARD LIBERA, Mathematical Sciences Department, University of Delaware, Newark, DE 19716. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Proposed problems (three copies) and solutions (two copies), both elementary and advanced to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, University of Calgary, Alberta, Canada T2N 1N4

Reprint Permission: A. B WILLCOX, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N W , Washington, DC 20036

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N W., Washington, DC 20036

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Back Issues: P. and H. BLISS CO., Middletown, CT 06457.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N W , Washington, DC 20036 and Montpelier, VT

The annual subscription price of $26 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $29 Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership) The nonmember/library subscription price is $90 per year

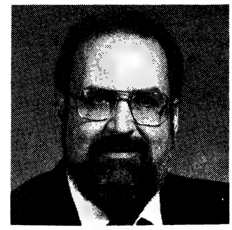
Second class postage paid at Washington, DC, and additional mailing offices

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Assocation of America, 1529 Eighteenth Street, N.W., Washington, DC 20077-9564.

PRINTED IN THE UNITED STATES OF AMERICA

An Algebraic Approach to Some Number-Theoretic Problems Arising from Paper-Folding Regular Polygons

JON FROEMKE AND JERROLD W. GROSSMAN



JON FROEMKE: I received my Ph.D. from the University of California at Berkeley in 1967 in universal algebra under the direction of Alfred Foster. I have been at Oakland University ever since. This article represents a return to number theory, my first love, and in an unusual way, a continuation of my interest in the arithmetic of primal algebras.

JERROLD W. GROSSMAN: I specialized in mathematical logic at Stanford University (B.S., M.S. in 1970), studied algebraic topology at M.I.T. (Ph.D. in 1974 under Daniel Kan), and have been at Oakland University since 1974. My research articles have dealt with algebraic topology, algebra, graph theory, combinatorics, probability and statistics, computer science, and now, apparently, number theory.

Introduction. Since 1983 Peter Hilton and Jean Pedersen have been studying the ramifications of an ingenious paper-folding construction they devised for approximating angles and regular polygons **[2]**, **[3]**, **[4]**, **[5]**, **[6]**, **[7]**. They are quickly led into number-theoretic questions. In this article we will see that with the use of some additional tools of graph theory, number theory, and algebra, we can more fully explain the phenomena they are studying. We also pose some new questions they have not raised and answer many of them. Perhaps not surprisingly, our investigations touch on several aspects of elementary number theory and lead quickly to some of its famous unsolved problems. On the other hand, Hilton and Pedersen pursue other interesting number-theoretic aspects of the construction, which we do not consider, so in no sense does our paper supersede their work.

This article is somewhat self-contained, in the sense that most of the number theory we need is reviewed herein. The underlying problems are extremely easy to state, and thus the article should be accessible to a wide audience. Indeed, the basic numerical construction proved fascinating to the first author's ten-year-old son, who could appreciate the questions being raised, but not, of course, the reasons behind their answers.

In what follows, we will let $\mathbf{Z}$ denote the set of integers, and for integers x and y we will let (x, y) and $[x, y]$ denote their greatest common divisor and least common multiple, respectively (with the obvious extensions to more than two variables).

Since the statement of the problems we are considering cannot be given until Section 2, it would be meaningless to give an outline of the paper at this point. The usual preview, section by section, is instead given at the end of Section 2.

1. Folding angles: the geometric motivation. Hilton and Pedersen have a clever method for approximating rational angles, to any desired degree of accuracy, using only an elementary paper-folding operation: bisecting an angle. (By rational angle we mean an angle whose measure is a rational number of degrees.) We summarize their procedure briefly. Although the geometric problem has no direct bearing on the rest of this paper, it serves as the motivation for considering the numerical construction of Section 2.

Let $r_0 = a\pi/b$ be a given acute rational angle, i.e., assume that $0 < a < b/2$, where a and b are integers with $(a, b) = 1$. Call b the *denominator* of r_0. For their purposes, Hilton and Pedersen assume initially that b is odd, and we will also (for now) make this assumption. Suppose angle r_0 is formed between the bottom edge of a long strip of paper and a fold in the paper (see FIGURE 1).

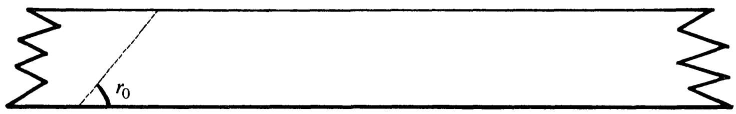


FIG. 1. Angle r_0 on a strip of paper.

If a is even, say $a = 2a'$, then by folding the original crease onto the bottom edge of the strip, we can construct angle $r_1 = a'\pi/b$ (see FIGURE 2).

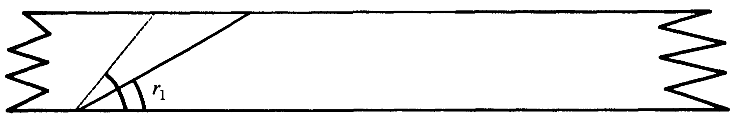


FIG. 2. Bisecting $a\pi/b$ when a is even.

On the other hand, if a is odd, then the supplementary adjacent interior angle at the top of the strip has measure $(b - a)\pi/b$, with $(b - a)$ even, say equal to $2a''$. We can, therefore, bisect it by folding the original crease onto the top edge of the strip and, thereby, form $r_1 = a''\pi/b$ along the top of the strip (see FIGURE 3).

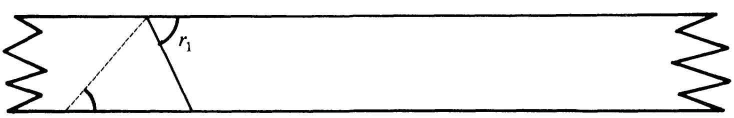


FIG. 3. Bisecting $(b - a)\pi/b$ when a is odd.

Iterating this procedure, we obtain a sequence of acute rational angles $r_0, r_1, r_2, \ldots$ with denominator b. It turns out (as a consequence of some observations in Section

2) that the sequence is cyclic, and hence that $r_0 = r_k$ for some k (to be called the quasiorder of 2 mod b). See FIGURE 4.

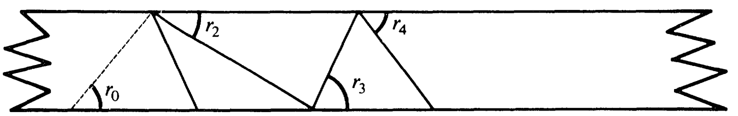


FIG. 4. $r_0 = r_4$.

Now if in fact r_0 is only an approximation to $a\pi/b$, say with a small error ε, then it is easy to see that the r_k constructed by this procedure is again an approximation to $a\pi/b$, but with error $\varepsilon/2^k$. Thus, by iterating the process, we can construct arbitrarily good approximations to $a\pi/b$.

Using this paper-folding procedure, Hilton and Pedersen devised a systematic method for constructing arbitrarily good approximations to any regular convex polygon or regular star polygon.

2. Charm bracelets: the numerical construction. Isolating the numerical ingredients of the foregoing geometric construction, we have simply the following operation on the set of positive integers less than b and relatively prime to b, where b is a fixed odd integer greater than 2:

$$\text{if } a \text{ is even, then halve it, i.e., let } a' = a/2;$$

$$\text{if } a \text{ is odd, then subtract it from } b, \text{ i.e., let } a' = b - a.$$

(The reader may have noticed the similarity of this operation to the one that appears in the Collatz problem **[9]**, although we have found nothing here that seems relevant to that notorious problem.) We will find it easier (but equivalent) in what follows to work instead with the opposite operation, which we now define.

DEFINITION. Let $V(b) = \{a \in \mathbf{Z} \mid 0 < a < b \text{ and } (a, b) = 1\}$. The function f_2: $V(b) \to V(b)$ is given by the rule

$$f_2(a) = \begin{cases} 2a & \text{if } a < b/2 \\ b - a & \text{if } a > b/2 \end{cases}.$$

(The subscript is 2 because of the multiplication by 2; we generalize this later.) To get a picture of this operation, we construct, for each odd $b > 1$, a graph whose vertices are the elements of $V(b)$ and whose edges are all the unordered pairs $(a, f_2(a))$. (If $b = 3$, then we put two edges between 1 and 2, since $f_2(1) = 2$ and $f_2(2) = 1$.) The graph for $b = 17$ is shown in FIGURE 5.

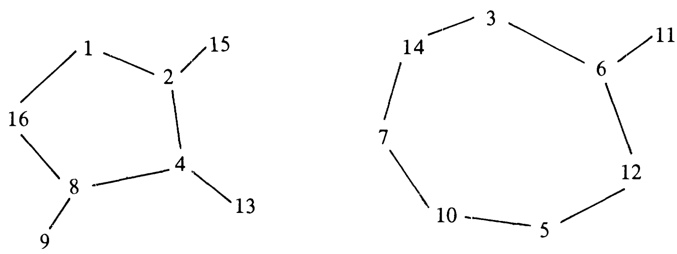


FIG. 5. The graph for $b = 17$.

The following observations about these graphs follow immediately from the definitions.

1. Each vertex a for which a is odd and greater than $b/2$ has degree 1, being adjacent only to $b - a$. Such values of a we will call *charms*.
2. Each vertex a for which a is even and less than $b/2$ has degree 3, being adjacent to $2a$, $a/2$, and the charm $b - a$.
3. Each vertex a for which a is even and greater than $b/2$ has degree 2, being adjacent to $b - a$ and $a/2$.
4. Each vertex a for which a is odd and less than $b/2$ has degree 2, being adjacent to $b - a$ and $2a$.

Now if we momentarily discard the charms, then every vertex has degree 2, and hence (by a trivial result of graph theory) the graph consists of one or more disjoint cycles (polygons). We call such a cycle, together with all the charms connected to it, a *bracelet*. We call the number of vertices in a bracelet its *weight*, and (continuing the jewelry metaphor) we call the number of noncharms in a bracelet its *size*. Finally we let $B(2, b)$ denote the number of bracelets in the graph. For $b = 17$, we see from FIGURE 5 that $B(2, 17) = 2$, that the weight of each bracelet is 8, and that the sizes of the two bracelets are 5 and 7, respectively.

Our bracelet graph replaces Hilton and Pedersen's "symbol" for representing the same information. They would display the second bracelet in FIGURE 5, for example, as

$$17\left|\begin{matrix} 3 & 7 & 5 \\ 1 & 1 & 2 \end{matrix}\right|.$$

This symbol is interpreted as follows: beginning with $a = 3$, we subtract a from 17 and divide the answer (14) as many times as we can by the number 2 (in this case, we divide once, obtaining the number 7 as the next value of a). The number of factors of 2 that were divided out appears in the second row, and the next value of a appears in the first row of the next column. The process is repeated until the original value of a is obtained. Two differences between our approaches are worth noting. First, Hilton and Pedersen list only the odd values of a less than $b/2$; the missing intermediate even values are implied and the second row in each column shows the number of missing even values. (Hilton and Pedersen use their symbol,

among other things, to succinctly encode the instructions for paper-folding approximations to star polygons.) Second, the symbol, as read from left to right, represents divisions by 2, rather than multiplications by 2, as in our approach, although Hilton and Pedersen also present a multiplication-based, as well as a division-based, "quasi-order algorithm." At this point, the differences are minor, but we will see that the bracelet symbolism clarifies the situation when we generalize, in Section 3, from multiplication (or division) by 2 to multiplication (or division) by an arbitrary number t.

We will sometimes find it convenient to modify our way of looking at things by identifying each number a in $V(b)$ with $b - a$. This is more in the spirit of the geometric construction, where we considered only acute angles.

DEFINITION. Let $D(b) = \{a \in \mathbf{Z} \mid 0 < a < b/2 \text{ and } (a, b) = 1\}$. The function g_2: $D(b) \to D(b)$ is given by the rule

$$g_2(a) = \begin{cases} 2a & \text{if } 2a < b/2 \\ b - 2a & \text{if } 2a > b/2 \end{cases}.$$

To obtain a visual model of the structure induced by g_2, we need only draw a box around each pair $(a, b - a)$ in the graphs obtained above (see FIGURE 6 for the situation when $b = 17$).

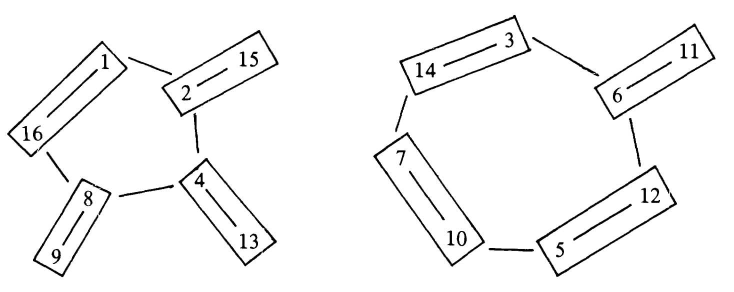


FIG. 6. Dominoes for $b = 17$.

We will call each such pair a *domino*, so that each bracelet can now be thought of simply as a cycle of dominoes. The cyclic geometric construction of Section 1 is represented by this graph: the domino containing $a \in D(b)$ corresponds to the acute angle $a\pi/b$. Note that as a consequence of this view, we obtain our first easy fact about bracelets.

PROPOSITION 1. *The weight of a bracelet is always even.*

If one computes (either by hand or, more easily, with a computer) the bracelet structures for some small values of b, especially for b prime, then one is led to an almost inexhaustible set of questions and conjectures. (To provide the reader with some examples for our discussion, we display a brief set of data in the Appendix.)

For example, do all the bracelets for a given b have the same weight? Do they all have sizes of the same parity? Is there any way to "compute" the number of bracelets, $B(2, b)$?

In the remainder of this article, then, we will explore such questions, with an emphasis on what can be said about the function B. In Section 3, we lay out the basic number-theoretic and group-theoretic way of looking at this problem, generalizing from 2 as the generator to an arbitrary t relatively prime to an arbitrary $b > 2$. In Section 4 we give a formula for computing the number of bracelets, their weights, and the parity of their sizes. Restricting ourselves to prime b in Section 5, we look more closely at how t and b determine the bracelet structure and specifically the number of bracelets. In Section 6 we extend the investigation to powers of primes and in Section 7 to arbitrary composite b; we will see that the set of prime factors of b tells essentially the whole story. Finally we close, in Section 8, with a few open-ended musings.

In addition to the work of Hilton and Pedersen, we mention that of J. B. Roberts [**13**], whose work anticipates some of ours, and H. P. Lawther, Jr. [**10**], who applied some related ideas to the splicing of telephone cables.

3. Some number theory and a slight generalization. In order to continue our investigation, it will be helpful to recall some elementary terminology and results from number theory and group theory.

Fix a (not necessarily odd) number $b > 2$. First note that we need not restrict ourselves to working with the set $V(b)$ of numbers relatively prime to b *and less than* b, because we are really working modulo b. Indeed, $V(b)$ is simply a collection of representatives for the set $\mathbf{Z}_b^*$ of reduced residue classes modulo b, which forms a finite abelian group under multiplication. Recall that the order of $\mathbf{Z}_b^*$, i.e., the cardinality of $V(b)$, is denoted by $\phi(b)$.

For t relatively prime to b, we consider the sequence $t, t^2, t^3, \dots$. For some $n > 0$ we must have $t^n \equiv 1 \pmod b$; indeed, $t^{\phi(b)} \equiv 1 \pmod b$. The least such n is called the *order of* t mod b, and we shall denote it by $o(t, b)$. In fact, $t^n \equiv 1$ $(\mathrm{mod}\ b)$ if and only if $o(t, b)$ is a divisor of n. Now it might happen that in the sequence $t, t^2, t^3, \dots,$ the number $-1 \pmod b$ occurs before 1 does; this possibility has fundamental significance for the problems we are studying. Thus we make the following definition (which appears in [**6**] and, under different names, in [**10**] and [**13**], as well).

DEFINITION. Let t and b be relatively prime integers, $b > 2$. The *quasi-order* of t mod b, denoted $q(t, b)$, is the least positive integer k such that $t^k \equiv \pm 1 \pmod b$. If $t^{q(t,b)} \equiv -1 \pmod b$, then we call t *basic* $(\mathrm{mod}\ b)$, and if $t^{q(t,b)} \equiv 1 \pmod b$, then we call t *nonbasic* $(\mathrm{mod}\ b)$.

Note that $q(t, b)$ must equal either $o(t, b)$ or $o(t, b)/2$, since if $t^k \equiv \pm 1$ $(\mathrm{mod}\ b)$, then $t^{2k} \equiv 1 \pmod b$. The following fact about quasi-order can be proved in a manner similar to the corresponding fact about order.

PROPOSITION 2. *Let t and b be relatively prime integers, $b > 2$. Then $t^n \equiv \pm 1 \pmod{b}$ if and only if $q(t, b)$ is a divisor of n.*

If $o(t, b) = \phi(b)$, then the set $\{t, t^2, t^3, \ldots, t^{\phi(b)}\}$ represents all the reduced residue classes mod b, and hence t generates $\mathbf{Z}_b^*$ under multiplication. In other words, $\mathbf{Z}_b^*$ is cyclic in this case. A well-known result of number theory, which we exploit in Section 5, states that $\mathbf{Z}_b^*$ is cyclic if and only if b equals 2, 4, p^n or $2p^n$ for p an odd prime and n a positive integer. In any case, the powers of $t \pmod{b}$ form a subgroup $\langle t \rangle$ of $\mathbf{Z}_b^*$.

In what follows we need to look at the subgroup $\langle t, -1 \rangle$ of $\mathbf{Z}_b^*$ generated by t and -1. If t is basic, then $-1 \in \langle t \rangle$, so in this case $\langle t, -1 \rangle = \langle t \rangle$ (hence our choice of the term *basic*). If t is nonbasic, then $\langle t, -1 \rangle$ contains all the elements of $\langle t \rangle$ together with their negatives. Thus in either case (but for different reasons), we obtain the following simple result.

PROPOSITION 3. *Let t and b be relatively prime integers, $b > 2$. Then $\{\pm t, \pm t^2, \pm t^3, \ldots, \pm t^{q(t,b)}\}$ is a set of representatives for $\langle t, -1 \rangle$, and the cardinality of $\langle t, -1 \rangle$ is $2q(t, b)$.*

Returning to the terminology of Section 2, we see that the bracelet containing 1 is exactly a collection of representatives for $\langle 2, -1 \rangle$. Furthermore, we can construct bracelets for values of t other than 2 simply by generalizing the definition of f_2 (and that of g_2), reducing all calculations modulo b. (Unfortunately, it no longer seems to be easy to characterize the charms.)

DEFINITION. Let t and b be relatively prime integers, $b > 2$. The function f_t: $V(b) \to V(b)$ and the function g_t: $D(b) \to D(b)$ are given by the following rules:

$$f_t(a) = \begin{cases} ta \bmod b & \text{if } a < b/2 \\ b - a & \text{if } a > b/2 \end{cases}$$

and

$$g_t(a) = \begin{cases} ta \bmod b & \text{if } ta \bmod b < b/2 \\ b - ta \bmod b & \text{if } ta \bmod b > b/2 \end{cases}.$$

As in the case $t = 2$, we let $B(t, b)$ be the number of bracelets formed in the construction. Note that Proposition 1 remains valid. As an example, we have the bracelet in FIGURE 7 for $b = 17$ and $t = 7$; thus $B(7,17) = 1$ and $q(7,17) = 8$.

Again, our function f_t (or g_t) and the bracelet graph replace the "quasi-order algorithm" and "symbol" of Hilton and Pedersen. Since all that is involved is multiplication modulo b, we avoid the cumbersome calculations and bookkeeping that they encounter when using the division-based approach when $t > 2$. On the other hand, their symbol does immediately determine the quasi-order and a criterion for whether t is basic (mod b).

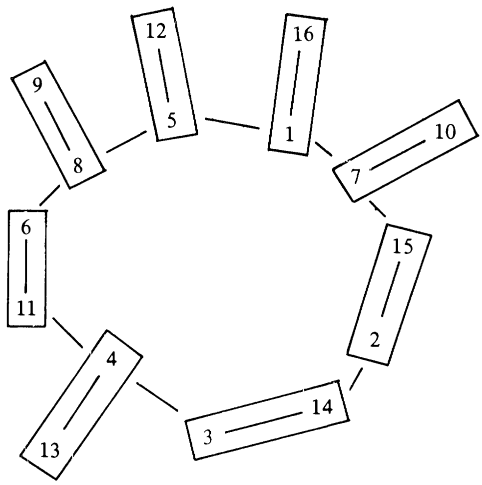


FIG. 7. Bracelet (with domino structure) for $b = 17$, $t = 7$.

4. A little group theory sheds some light. In Section 3 we saw that the bracelet containing 1 is just (a collection of representatives for) the subgroup $\langle t, -1\rangle$. In this section we apply some rudimentary group theory to discover what the remaining bracelets are, what the number of bracelets signifies, and how the domino structure can be interpreted. To avoid the awkward construction in parentheses in the first sentence of this paragraph, we will identify elements of $\mathbf{Z}_b^*$ with their representatives and think of the bracelets as actually containing elements of $\mathbf{Z}_b^*$.

THEOREM 1. *The bracelets for a given $b > 2$ and t relatively prime to b are precisely the cosets of $\langle t, -1\rangle$ in $\mathbf{Z}_b^*$. They all have weight $2q(t, b)$, and the number of bracelets is given by*

$$B(t, b) = \frac{\phi(b)}{2q(t, b)} = \textit{index of } \langle t, -1\rangle \text{ in } \mathbf{Z}_b^*.$$

Proof. For each $a \in \mathbf{Z}_b^*$, the bracelet containing a consists of all numbers $(\bmod b)$ of the form at^i, $i = 1, 2, \ldots$. Thus it consists of precisely $\{\pm at, \pm at^2, \ldots, \pm at^{q(t,b)}\}$, which is the coset $a\langle t, -1\rangle$. The remaining statements follow from Proposition 3 and the definitions.

We next turn to an analysis of the sizes of the bracelets, i.e., the number of vertices in each bracelet which are not charms. We saw in FIGURE 5 that the sizes need not all be the same for a given b and t. On the other hand, the following theorem allows us to calculate their parity.

THEOREM 2. *The sizes of all the bracelets for a given $b > 2$ and t relatively prime to b have the same parity. This parity is even if t is basic $(\bmod b)$ and $q(t, b)$ is odd, or if t is nonbasic $(\bmod b)$ and $q(t, b)$ is even; and otherwise this parity is odd.*

Proof. Each step in a traversal of the cycle of a bracelet (i.e., the vertices which are not charms) corresponds either to multiplication by t (with a move to the next domino) or to multiplication by -1 (staying within the same domino). If s is the size of the bracelet, then after s steps (and not before), we will have returned to the starting point. If we performed k multiplications by t and $s-k$ multiplications by -1, then we must have (mod b)

$$at^k(-1)^{s-k} \equiv a$$

or simply

$$t^k \equiv (-1)^{s-k}.$$

Therefore, k must be $q(t, b)$ and $(-1)^{s-k}$ must be -1 or 1 according as whether t is basic or nonbasic (mod b), independent of the particular bracelet, i.e., independent of s. Thus the parity of $s-k$, and, hence, of s, depends only on b and t, in the manner stated. (It is also possible to give a proof based on a generalization of Gauss's Lemma due to Emma Lehmer [**11**].)

Since the size of a bracelet plus the number of its charms is even by Proposition 1, a statement similar to Theorem 2 holds as well for the number of charms.

Finally, we note that $D(b)$ can be viewed as a collection of representatives for the factor group $\mathbf{Z}_b^*/\langle -1\rangle$, which is essentially just the set of dominoes.

5. Some more number theory sheds some more light. In Section 4 we reduced questions about the sizes, weights, and numbers of bracelets to questions about $q(t, b)$ and about $t^{q(t,b)}$ mod b, i.e., whether t is basic or nonbasic (mod b). We now continue this study by classifying the bracelet structure into one of eight types, defined in terms of the sign of $t^{q(t,b)}$ mod b, the parity of $q(t, b)$, and the parity of $B(t, b)$. The following table shows the types, together with the smallest example in which $t = 2$ and b is prime (the dashes in the last column indicate that—as Theorem 3 will state—these types are impossible for prime b).

TABLE 1. TYPES OF BRACELET STRUCTURES.

Type 1.	t basic	$q(t, b)$ odd	$B(t, b)$ odd	$(b = 3)$
Type 2.	t nonbasic	$q(t, b)$ odd	$B(t, b)$ odd	$(b = 7)$
Type 3.	t basic	$q(t, b)$ even	$B(t, b)$ odd	$(b = 5)$
Type 4.	t nonbasic	$q(t, b)$ even	$B(t, b)$ odd	—
Type 5a.	t basic	$q(t, b)$ odd	$B(t, b)$ even	$(b = 281)$
Type 5b.	t nonbasic	$q(t, b)$ odd	$B(t, b)$ even	$(b = 73)$
Type 5c.	t basic	$q(t, b)$ even	$B(t, b)$ even	$(b = 17)$
Type 5d.	t nonbasic	$q(t, b)$ even	$B(t, b)$ even	—

By using quadratic residues and the fact that $\mathbf{Z}_b^*$ is cyclic for certain values of b (including all odd prime numbers), we will now determine, for such values of b, when each of these cases occurs. Recall that an integer x is a *quadratic* residue mod b if $x \equiv y^2$ (mod b) for some integer y.

THEOREM 3. *Suppose that* $\mathbf{Z}_b^*$ *is cyclic,* $b \neq 2, 4$ (*i.e.,* $b = p^n$ *or* $2p^n$, *for* p *an odd prime and* n *a positive integer*), *and suppose that* $(t, b) = 1$. *Then*

$$B(t, b) = \frac{p^{n-1}(p-1)}{2q(t, b)}.$$

Furthermore,

a) *if neither* t *nor* -1 *are quadratic residues* $\bmod b$, *then* $q(t, b)$ *is odd,* $t^{q(t,b)} \equiv -1 \pmod b$, $\langle t \rangle = \langle t, -1 \rangle \neq \langle -t \rangle$, *and* $B(t, b)$ *is odd* (*Type* 1);

b) *if* t *is a quadratic residue* $\bmod b$ *but* -1 *is not, then* $q(t, b)$ *is odd,* $t^{q(t,b)} \equiv 1 \pmod b$, $\langle t \rangle \neq \langle t, -1 \rangle = \langle -t \rangle$, *and* $B(t, b)$ *is odd* (*Type* 2);

c) *if* -1 *is a quadratic residue* $\bmod b$ *but* t *is not, then* $q(t, b)$ *is even,* $t^{q(t,b)} \equiv -1 \pmod b$, $\langle t \rangle = \langle t, -1 \rangle = \langle -t \rangle$, *and* $B(t, b)$ *is odd* (*Type* 3);

d) *if both* t *and* -1 *are quadratic residues* $\bmod b$, *then* $B(t, b)$ *is even* (*Type* 5); *if Type* 5*a occurs, then* $\langle t \rangle = \langle t, -1 \rangle \neq \langle -t \rangle$; *if Type* 5*b occurs, then* $\langle t \rangle \neq \langle t, -1 \rangle = \langle -t \rangle$; *and if Type* 5*c occurs, then* $\langle t \rangle = \langle t, -1 \rangle = \langle -t \rangle$;

e) *Type* 4 *and Type* 5*d cannot occur.*

Proof. Since $b = p^n$ or $2p^n$, it is easy to compute the order of $\mathbf{Z}_b^*$, namely $\phi(b) = p^{n-1}(p-1)$. Thus the displayed equation follows from Theorem 1.

a) Let the integer g be a generator of $\mathbf{Z}_b^*$, and write $t \equiv g^r \pmod b$. Since t is assumed not to be a quadratic residue mod b, r must be odd. Thus $t^{\phi(b)/2} \equiv (-1)^r \equiv -1 \pmod b$. We claim that, therefore, $q(t, b)$, which must be a divisor of $\phi(b)/2$ by Proposition 2, is odd. Indeed, since -1, which is congruent to $g^{\phi(b)/2}$, is assumed not to be a quadratic residue mod b, it must be that $\phi(b)/2$ is odd. It follows, then, that $t^{q(t,b)} \equiv -1 \pmod b$, i.e., t is basic. Combining all this we see that $B(t, b)$ is odd. Finally, since t is basic, it is clear that t generates $\langle t, -1 \rangle$; on the other hand, since $q(t, b)$ is odd, $(-t)^{q(t,b)} \equiv 1 \pmod b$, so $-t$ does not generate $\langle t, -1 \rangle$.

b) The proof is similar to part (a), except that we use $-t$ in place of t. This works because, given that t is a quadratic residue and -1 is not, $-t$ is not a quadratic residue.

c) The proof is similar to part (a), except that since $\phi(b)/2$ is even (by the assumption that -1 is a quadratic residue), $q(t, b)$ must absorb all the factors of 2 in $\phi(b)/2$ in order for $t^{\phi(b)/2}$ to be congruent to -1. Thus $q(t, b)$ is even, and, again, $B(t, b)$ is odd. Since t is basic, $\langle t \rangle = \langle t, -1 \rangle$; and since, in addition, $q(t, b)$ is even, $\langle -t \rangle = \langle t, -1 \rangle$, as well.

d) If $q(t, b)$ is odd, then since $\phi(b)/2$ is even, $B(t, b)$ is even, and the other statements follow as above. If $q(t, b)$ is even, then $t^{q(t,b)} \equiv -1 \pmod b$, since otherwise $t^{q(t,b)/2}$ would also be congruent to $\pm 1 \pmod b$, contradicting the definition of $q(t, b)$. (This last statement uses the fact that the number 1 has exactly two square roots in $\mathbf{Z}_b^*$ [**14**, p. 247].) Thus the order of t is $2q(t, b)$. It follows that $\langle t \rangle = \langle t, -1 \rangle = \langle -t \rangle$. Now since t is a quadratic residue, $t^{\phi(b)/2} \equiv 1 \pmod b$. Therefore $2q(t, b)$ divides $\phi(b)/2$, and so 2 divides $\phi(b)/(2q(t, b)) = B(t, b)$.

e) If $q(t, b)$ were even and $t^{q(t,b)} \equiv 1 \pmod b$, then $t^{q(t,b)/2} \equiv \pm 1 \pmod b$, as in

the proof of (d). This contradicts the definition of $q(t, b)$, so Type 4 and Type 5d cannot occur. (We will see later that they can occur when $\mathbf{Z}_b^*$ is not cyclic.)

We note the following corollary of Theorems 2 and 3.

COROLLARY 1. *Suppose that $\mathbf{Z}_b^*$ is cyclic, $b \neq 2, 4$ (i.e., $b = p^n$ or $2p^n$, for p an odd prime and n a positive integer), and suppose that $(t, b) = 1$. Then there are an even number of bracelets if and only if both t and -1 are quadratic residues* $\bmod b$. *Furthermore, the sizes of the bracelets are even if neither t nor -1 are quadratic residues, and odd if exactly one of t and -1 is a quadratic residue (the sizes may be either even or odd if both are quadratic residues).*

The determination of the quadratic residue status of 2 and $-1 \bmod b$ is always very simple in the cases we are considering. The following proposition is a standard exercise in number theory (see [**14**, pp. 254 and 256] for the flavor of the arguments involved).

PROPOSITION 4. *If $\mathbf{Z}_b^*$ is cyclic, $b \neq 2$ or 4, then -1 is a quadratic residue* $\bmod b$ *if and only if $b \equiv 1$ or 2 (mod 4). If $b = p^n$ for p an odd prime and n a positive integer, then 2 is a quadratic residue* $\bmod b$ *if and only if $p \equiv 1$ or 7 (mod 8).*

Combining Proposition 4 with Theorem 3, we see that, for $t = 2$ and b prime, types 1, 2, 3, and 5 occur when b is congruent to 3, 7, 5, and 1 (mod 8), respectively. Since by Dirichlet's Theorem [**14**, p. 375] there are an infinite number of primes in each of these congruence classes, we see that each of these types occurs infinitely often. In particular, *for $t = 2$ there are at least two bracelets for infinitely many values of b.* In case t is some number other than 2, results similar to Proposition 4 can be obtained using the Law of Quadratic Reciprocity, and again we will have an infinite number of occurrences of each of types 1, 2, 3, and 5. We leave it as an exercise to show that the residue class of $b \bmod 12$ determines the type when b is prime and $t = 3$.

Two obvious questions about the number of bracelets arise when one looks at the data in the Appendix: is there only one bracelet infinitely often, and can there be an arbitrarily large number of bracelets? In the remainder of this section, we will consider these questions for prime b.

The answer to both questions is an easy "yes" if we are willing to think of B as a function with two arguments, rather than thinking of t as a fixed parameter.

THEOREM 4. *For infinitely many pairs (t, b) with b prime, $B(t, b) = 1$. As a function of two variables, $B(t, b)$ is unbounded, even with b restricted to being prime.*

Proof. For the first claim, we need only observe that for prime b, we can choose the integer t to be a generator of the cyclic group $\mathbf{Z}_b^*$. For the second, we can make the number of bracelets arbitrarily large simply by taking t to be $b - 1$, since then the quasi-order of t will be 1, and so the number of bracelets will be $(b - 1)/2$ by Theorem 1.

The questions become much harder if we fix t (still insisting that b be prime). By Theorem 3, $B(2, b) = 1$ if and only if the integer 2 (or in some cases -2) is a generator of $\mathbf{Z}_b^*$. Number theorists call 2 (or -2) a *primitive root* in this case. It is conjectured, but not known, that 2 is a primitive root for infinitely many primes. If this conjecture were true (in a somewhat stronger form, since we would need to restrict ourselves to Type 1 or Type 3), then we could conclude that there is only one bracelet infinitely often. Similar statements could be made for other values of t.

Another long-standing conjecture is that there are an infinite number of primes p for which $b = 2p + 1$ is also prime [**15**, p. 129]. If p and b are such primes, then it is easy to show that $q(t, b) = p$, and therefore that $B(t, b) = 1$, for all t not congruent to $\pm 1 \pmod b$. In particular, if the conjecture is true, then for each $t > 1$ we would have $B(t, b) = 1$ infinitely often.

At the other end of the spectrum, $B(t, b)$ is large when $q(t, b)$ is small compared to b. The smallest $q(t, b)$ occurs when $t^{q(t,b)} = b \pm 1$, i.e., when $b = t^{q(t,b)} + 1$ or $b = t^{q(t,b)} - 1$. (There is currently a great deal of interest in the prime factorization of $t^k \pm 1$ for small t and large k; see [**1**].) In the particular case of $t = 2$, the questions of whether $2^k \pm 1$ are prime are well known. Prime numbers of the form $2^k + 1$ are the *Fermat primes*, and only five of them are known: 3, 5, 17, 257, and 65537. Thus, for example, $B(2{,}65537) = 65536/32 = 2048$. Prime numbers of the form $2^k - 1$ are the *Mersenne primes*, and only about 30 are known. (The largest Mersenne prime currently known is $2^{216091} - 1$.) Thus, for example, there are 315 bracelets when b is the Mersenne prime 8191. It is not known whether there are an infinite number of Fermat or Mersenne primes. If there are, then clearly $B(2, b)$ is unbounded for prime b. We can still derive this conclusion, however, by looking at the prime factors of the Fermat numbers.

THEOREM 5. *Let t be a fixed integer greater than* 1. *Then $B(t, b)$ is unbounded for prime numbers b.*

Proof. We use the following well-known lemma (whose proof is not hard—see [**15**, p. 343], for example): *If p is an odd prime divisor of $t^{2^n} + 1$, then $p = 2^{n+1}i + 1$ for some positive integer i.* Now for any positive integer M we can guarantee that $2^{n+1}i + 1$ is not prime for $1 \leqslant i \leqslant M$ by taking $n = \prod_{i=1}^{M}\phi(2i + 1)$. Indeed, since $2^{\phi(2i+1)} \equiv 1 \pmod{2i + 1}$, we have $2^n \equiv 1 \pmod{2i + 1}$, and, therefore, $2^{n+1}i + 1 \equiv 2i + 1 \equiv 0 \pmod{2i + 1}$. Thus we can take b to be an odd prime divisor of $t^{2^n} + 1$, which must exist since even if $t^{2^n} + 1$ is not odd, it is congruent to 2 (mod 4). By the lemma and the choice of n, we know that $b = 2^{n+1}i + 1$ for some $i > M$. On the other hand, the quasi-order of t is at most 2^n, since $t^{2^n} \equiv -1 \pmod b$, so by Theorem 3, $B(t, b)$ is at least $(b - 1)/2^{n+1} = i > M$.

6. Powers of primes. In the last section we looked at the quasi-order $q(t, b)$ and the number of bracelets $B(t, b)$, especially for odd prime b. We now make some further progress in the case in which b is a prime power. Of course Theorem 3 applies in the case of powers of odd primes. Here we want to relate $q(t, b)$ to $q(t, p)$ and $B(t, b)$ to $B(t, p)$ when $b = p^n$ where p is an odd prime, and also to compute $q(t, b)$ and $B(t, b)$ when $b = 2^n$.

We first state without proof a fundamental result relating the order of t mod b to the order of t mod pb [**12**, p. 364]. We need to assume at least that $t \neq \pm 1$ here, but for simplicity we will assume that $t > 1$.

PROPOSITION 5. *Let $b = p^n$ and $b' = p^{n+1}$, where p is an odd prime and $n \geqslant 1$. Assume $t > 1$ and $(t, b) = 1$. Then $o(t, b')$ is equal to either $o(t, b)$ or $po(t, b)$, and the second possibility holds precisely on the set of all $n \geqslant N(t, p)$ for some $N(t, p) \geqslant 1$ depending on t and p.*

In other words, as n increases, the order of t mod p^n eventually increases by the same factor as p^n increases, i.e., eventually $\phi(b)/o(t, b)$ becomes constant as n increases. In fact, it seems to be only rarely that $\phi(b)/o(t, b)$ increases at all, even as b goes from p to p^2. Primes p for which this does occur seem not to have been given a name in the number theory literature, so we will call them *Wieferich primes*, after an early twentieth-century mathematician who studied them in connection with Fermat's Last Theorem [**16**].

DEFINITION. A prime p is called *Wieferich* with respect to t if the order of t mod p^2 is the same as the order of t mod p.

The only Wieferich primes with respect to $t = 2$, less than 31 million, are 1093 and 3511 ([**8**] gives the complete—and very small—table of primes known (in 1965) to be Wieferich with respect to prime $t \leqslant 43$).

We now show that the behavior of the order of t mod p^n as n increases extends also to the behavior of the quasi-order, and hence that the number of bracelets for $b = p^n$ becomes constant for large enough n.

THEOREM 6. *Let $b = p^n$ and $b' = p^{n+1}$, where p is an odd prime and $n \geqslant 1$. Assume $t > 1$ and $(t, b) = 1$. Then*

$$\frac{q(t, b')}{q(t, b)} = \frac{o(t, b')}{o(t, b)} = \begin{cases} 1 & \text{if } n < N(t, p) \\ p & \text{if } n \geqslant N(t, p) \end{cases},$$

where $N(t, p)$ is the number guaranteed by Proposition 5.

Proof. The final equality follows from Proposition 5. As for the first, because of the remark above Proposition 2, there are two cases to consider.

i) Suppose $q(t, b) = o(t, b) = k$, so $t^k \equiv 1 \pmod{b}$. Hence k must be odd (otherwise $t^{k/2} \equiv \pm 1 \pmod{b}$, as in the proof of Theorem 3.d). By Proposition 5, $o(t, b')$ is also odd. Therefore, since $q(t, b')$ cannot be $o(t, b')/2$, it must equal $o(t, b')$, and the desired equality follows.
ii) Suppose $q(t, b) = o(t, b)/2 = k$. Thus $o(t, b) = 2k$, and so $o(t, b')$ is also even by Proposition 5. Therefore by the same reasoning as in (i), $q(t, b')$ must be $o(t, b')/2$ and not $o(t, b')$, and the desired equality follows.

COROLLARY 2. *Let p be an odd prime and $t > 1$ a number not divisible by p. If p is not Wieferich with respect to t, then*

$$B(t, p^n) = B(t, p).$$

In any case,

$$B(t, p^n) = p^{\min(n, N(t,p))-1} B(t, p),$$

where $N(t, p)$ is the number guaranteed by Proposition 5.

In other words, the number of bracelets for a prime power is eventually (and nearly always, it seems) independent of the power.

We now turn briefly to the case $b = 2^n$, with $n \geqslant 2$. Thus $\mathbf{Z}_b^*$ is not cyclic (unless $b = 4$), but it turns out [**14**, p. 205] that $\mathbf{Z}_b^*/\langle -1\rangle$ is. In contrast to Theorem 3, here usually neither t nor $-t$ generates the bracelet containing 1 (i.e., the subgroup $\langle t, -1\rangle$). We summarize what happens in this case but omit the proof. In terms of the classification given in TABLE 1, all types are possible here except Type 3 and Type 5c; almost always it is Type 4 or Type 5d.

THEOREM 7. *Let $b = 2^n$, $n \geqslant 2$, and let t be an odd integer. Then*

$$B(t, b) = \left(\frac{t \pm 1}{4}, \frac{b}{4}\right),$$

where the sign is chosen so that $(t \pm 1)/4$ is an integer. Also,

$$q(t, b) = \frac{2^{n-2}}{B(t, b)}.$$

Furthermore, t is nonbasic $\operatorname{mod} b$ unless $t \equiv -1 \pmod b$.

7. The general composite case. We look finally at what can be said about $q(t, b)$ and $B(t, b)$ for b composite.

For simplicity, we will carry out the analysis in the case $b = p_1 p_2$, where p_1 and p_2 are distinct odd primes, but the argument will generalize to yield Theorems 8 and 9 below. Let $(t, b) = 1$, and let r_1 and r_2 be the quasi-orders of $t \bmod p_1$ and mod p_2, respectively. We will need to compare the highest powers of 2 dividing r_1 and r_2, so we write $r_1 = 2^{x_1} r_1'$ and $r_2 = 2^{x_2} r_2'$, where r_1' and r_2' are odd. Our goal is to determine $q(t, b)$ and $B(t, b)$ in terms of $q(t, p_i)$ and $B(t, p_i)$, $i = 1, 2$.

We first recall what is essentially the uniqueness part of the Chinese Remainder Theorem.

PROPOSITION 6. *If $a \equiv b \pmod{m_i}$ for $i = 1, 2, \ldots, n$, and if the m_i are pairwise relatively prime, then $a \equiv b \pmod{m_1 m_2 \cdots m_n}$.*

Since $t^{2r_i} \equiv 1 \pmod{p_i}$, clearly $t^{2[r_1, r_2]} \equiv 1 \pmod{p_i}$. Hence $t^{2[r_1, r_2]} \equiv 1 \pmod b$ by Proposition 6. Therefore, $q(t, b)$ is a divisor of $2[r_1, r_2]$. On the other hand, since $t^{q(t,b)} \equiv \pm 1 \pmod b$, we have $t^{q(t,b)} \equiv \pm 1 \pmod{p_i}$ for $i = 1, 2$. Thus by Proposition 2, $q(t, b)$ is a multiple of both r_1 and r_2 and hence of their least common multiple. Combining these two statements, we see that either $q(t, b) = [r_1, r_2]$ or

$q(t, b) = 2[r_1, r_2]$. To determine which formula applies, we need to look at the four cases.

Case 1. $t^{r_i} \equiv 1 \pmod{p_i}$ for $i = 1, 2$. Then $t^{[r_1, r_2]} \equiv 1 \pmod{b}$ by Proposition 6, so $q(t, b) = [r_1, r_2]$.

Case 2. $t^{r_i} \equiv -1 \pmod{p_i}$ for $i = 1, 2$. If r_1 and r_2 have the same highest power of 2 as a factor, i.e., $x_1 = x_2$ in the notation established above, then $[r_1, r_2]/r_i$ is odd, so $t^{[r_1, r_2]} \equiv -1 \pmod{p_i}$, for $i = 1, 2$. Thus $t^{[r_1, r_2]} \equiv -1 \pmod{b}$ by Proposition 6, and hence $q(t, b) = [r_1, r_2]$. If, however, $x_1 \neq x_2$, then one of $[r_1, r_2]/r_1$ and $[r_1, r_2]/r_2$ is even and the other is odd, so $t^{[r_1, r_2]}$ has different values reduced modulo p_1 and modulo p_2. Therefore $t^{[r_1, r_2]}$ cannot be congruent to either 1 or -1 $\pmod{b}$. It follows that $q(t, b) = 2[r_1, r_2]$ in this subcase.

Case 3. $t^{r_1} \equiv -1 \pmod{p_1}$ and $t^{r_2} \equiv 1 \pmod{p_2}$. Using the same reasoning as in Case 2, we see that $q(t, b) = [r_1, r_2]$ if $x_1 < x_2$ and $q(t, b) = 2[r_1, r_2]$ if $x_1 \geqslant x_2$.

Case 4. $t^{r_1} \equiv 1 \pmod{p_1}$ and $t^{r_2} \equiv -1 \pmod{p_2}$. The situation is similar to Case 3.

As a corollary of this analysis, we see that, except for a factor of (r_1, r_2) and a possible factor of 2, the "number of bracelets" function is multiplicative. Indeed, since $B(t, b) = \phi(b)/(2q(t, b))$, we see that if $q(t, b) = [r_1, r_2]$, then

$$B(t, b) = \frac{\phi(b)}{2[r_1, r_2]} = \frac{(p_1 - 1)(p_2 - 1)}{2r_1r_2/(r_1, r_2)} = 2(r_1, r_2)B(t, p_1)B(t, p_2),$$

and if $q(t, b) = 2[r_1, r_2]$, then

$$B(t, b) = \frac{\phi(b)}{4[r_1, r_2]} = \frac{(p_1 - 1)(p_2 - 1)}{4r_1r_2/(r_1, r_2)} = (r_1, r_2)B(t, p_1)B(t, p_2).$$

As one example, let $t = 2$, $p_1 = 5$, and $p_2 = 7$. Then $r_1 = 2$, $r_2 = 3$, $t^{r_1} \equiv -1 \pmod{p_1}$, and $t^{r_2} \equiv 1 \pmod{p_2}$. Case 3 applies, with $x_1 > x_2$, so $q(2, 35) = 2[2, 3] = 12$, and $B(2, 35) = B(2, 5)B(2, 7) = 1$. For another example, suppose $t = 2$, $p_1 = 3$, and $p_2 = 11$. Then we are in Case 2 with $2^1 \equiv -1 \pmod{3}$ and $2^5 \equiv -1 \pmod{11}$, so $q(2, 33) = [1, 5] = 5$ and hence $B(2, 33) = 2B(2, 3)B(2, 11) = 2$.

Nothing in the foregoing discussion except the calculations of $B(t, b)$ required that p_1 and p_2 were odd primes, only that each was greater than 2 and $(p_1, p_2) = 1$. Thus, generalizing in a straightforward way to arbitrary b, we obtain the following theorems, which give formulas for computing quasi-orders and numbers of bracelets for composite numbers in terms of these invariants for their prime-power factors. (In Corollary 3 we see that the number of bracelets ultimately does not even depend on the powers.) All eight types (see TABLE 1) are possible. We leave verification of the details to the reader.

THEOREM 8. *Let* $b = \prod_{i=1}^{n} p_i^{e_i}$, *where the* p_i *are distinct primes dividing* b *and* $b \not\equiv 2 \pmod{4}$; *or let* $b = 2\prod_{i=1}^{n} p_i^{e_i}$, *where the* p_i *are distinct odd primes dividing* b.

Suppose $t > 1$ and $(t, b) = 1$. Let r_i be the quasi-order of t mod $p_i^{e_i}$, and for each i write $r_i = 2^{x_i} r_i'$, where r_i' is odd. Then

a) *if t is nonbasic $(\text{mod } p_i^{e_i})$ for all i, then t is nonbasic $(\text{mod } b)$, and $q(t, b) = [r_1, r_2, \ldots, r_n]$;*
b) *if t is basic $(\text{mod } p_i^{e_i})$ for all i, then*
   i) *if $x_1 = x_2 = \cdots = x_n$, then t is basic $(\text{mod } b)$ and $q(t, b) = [r_1, r_2, \ldots, r_n]$, and*
   ii) *otherwise, t is nonbasic $(\text{mod } b)$ and $q(t, b) = 2[r_1, r_2, \ldots, r_n]$;*
c) *if t is basic $(\text{mod } p_i^{e_i})$ for some i, and t is nonbasic $(\text{mod } p_j^{e_j})$ for some j, then t is nonbasic $(\text{mod } b)$, and*
   i) *if $\max(x_1, x_2, \ldots, x_n) = x_i$ for some i for which t is basic $(\text{mod } p_i^{e_i})$, then $q(t, b) = 2[r_1, r_2, \ldots, r_n]$, and*
   ii) *if $\max(x_1, x_2, \ldots, x_n) > x_i$ for every i for which t is basic $(\text{mod } p_i^{e_i})$, then $q(t, b) = [r_1, r_2, \ldots, r_n]$.*

THEOREM 9. *Let $b = \prod_{i=1}^n p_i^{e_i}$, where the p_i are distinct primes dividing b and $b \not\equiv 2 \pmod 4$; or let $b = 2\prod_{i=1}^n p_i^{e_i}$, where the p_i are distinct odd primes dividing b. Suppose $t > 1$ and $(t, b) = 1$. Let r_i be the quasi-order of t mod $p_i^{e_i}$. Then*

$$B(t, b) = 2^{n-1-\varepsilon} \frac{r_1 r_2 \cdots r_n}{[r_1, r_2, \ldots, r_n]} \prod_{i=1}^n B(t, p_i^{e_i}),$$

where ε is either 0 or 1, depending on which of the cases in Theorem 8 applies: $\varepsilon = 0$ in cases (a), (b.i), and (c.ii), and $\varepsilon = 1$ in the cases (b.ii) and (c.i).

Note that Theorem 9 also provides a simpler proof that $B(t, b)$ is unbounded as a function of (not necessarily prime) b, since we need only take n large to make $B(t,b)$ large.

Finally suppose $b = \prod_{i=1}^n p_i^{e_i}$, where the p_i are distinct primes dividing b, with $t > 1$ and $(t, b) = 1$. If the exponents e_i are large enough, then it is not hard to show by Theorems 5 and 6 and Corollary 2 that all the terms on the right-hand side of the displayed equation in Theorem 9 are independent of the exponents. Thus we have the following Corollary—which we find most remarkable—that *the number of bracelets for b is independent of the exponents in the prime factorization of b once the exponents get large enough.*

COROLLARY 3. *Let $p_1, p_2, \ldots, p_n$ be distinct primes, none of which divides $t > 1$. Then there exist constants $E_1, E_2, \ldots, E_n$ and C (depending on $p_1, p_2, \ldots, p_n$ and t) such that if $b = \prod_{i=1}^n p_i^{e_i}$ with $e_i \geq E_i$ for $i = 1, 2, \ldots, n$, then $B(t, b) = C$.*

As a comprehensive example, let us study $b = 720$, $t = 7$. Since 720 factors as $2^4 3^2 5$, we first apply Theorem 7 to obtain $B(7,16) = 2$, $q(7,16) = 2$, and $7^2 \equiv 1 \pmod{16}$. Then we apply Corollary 2 (since 3 is not Wieferich with respect to 7) to obtain $B(7,9) = B(7,3) = B(1,3) = 1$ and $q(7,9) = \phi(9)/2 = 3$; and we apply Theorem 3b (or calculate directly) to obtain $7^3 \equiv 1 \pmod 9$. Finally we compute that $B(7,5) = B(2,5) = 1$, $q(7,5) = q(2,5) = 2$, and $7^2 \equiv -1 \pmod 5$. If we com-

bine all of these by Theorem 8.c.i, then we get $q(7,720) = 2[2,3,2] = 12$ and $7^{12} \equiv 1 \pmod{720}$; and applying Theorem 9, we see that $B(7,720) = 2^{3-1-1}(2 \cdot 3 \cdot 2/[2,3,2]) \cdot 2 \cdot 1 \cdot 1 = 8$. On the other hand, to see the force of Corollary 3, we have that if $e_1 \geqslant 5$, $e_2 \geqslant 1$, and $e_3 \geqslant 2$, then $B(7, 2^{e_1}3^{e_2}5^{e_3}) = 80$, since Theorem 8.c.ii applies and 5 is Wieferich with respect to 7 (N(in fact 7, 5) $= 2$).

Added in proof: We note that Theorem 8 answers a question raised by Man Keung Siu **[17]**, who obtained for the case $t = 2$ a sharpening of a part of Theorem 3.

8. Afterword. We leave the reader with a comment, a generalization, and an open question.

In some sense our investigation (except for Section 7 and the end of Section 6) has focused on a structure which is extremely well understood: the cyclic group. Nevertheless, interesting questions emerged, and the progress was somewhat impeded by famous unsolved problems in elementary number theory, such as whether 2 is a primitive root for infinitely many primes. We found it fascinating that this problem touched so many corners of elementary number theory. It could serve as a fruitful area of "research" for an undergraduate following an introductory number theory course.

The quasi-order of $t \bmod b$ was defined to be the least positive integer k such that t^k represents an element of the subgroup $\{-1, 1\}$ of $\mathbf{Z}_b^*$. More generally, for other subgroups S of $\mathbf{Z}_b^*$, we could define the quasiorder of $t \bmod b$ relative to S. Does anything interesting emerge?

Looking at the Appendix, we note that the sizes of the bracelets for a given b tend to be roughly equal, and that the bracelet containing 1 is usually the bracelet with the smallest size. Are any theorems lurking here?

Appendix: Some data. Listed here are the number of bracelets and the weights and sizes of the bracelets, for $t = 2$, b odd, $3 \leqslant b \leqslant 99$. The size of the bracelet containing 1 is marked with * when there is more than one bracelet.

b	$B(2,b)$	weights	sizes
3	1	2	2
5	1	4	3
7	1	6	5
9	1	6	4
11	1	10	8
13	1	12	9
15	1	8	6
17	2	8, 8	5*, 7
19	1	18	14
21	1	12	8
23	1	22	17
25	1	20	15
27	1	18	14
29	1	28	21
31	3	10, 10, 10	7*, 7, 9
33	2	10, 10	6*, 8

b	$B(2, b)$	weights	sizes
35	1	24	18
37	1	36	27
39	1	24	18
41	2	20, 20	13*, 17
43	3	14, 14, 14	10*, 10, 12
45	1	24	18
47	1	46	35
49	1	42	31
51	2	16, 16	12*, 12
53	1	52	39
55	1	40	30
57	2	18, 18	12*, 14
59	1	58	44
61	1	60	45
63	3	12, 12, 12	8*, 10, 10
65	4	12, 12, 12, 12	7*, 9, 9, 11
67	1	66	51
69	1	44	32
71	1	70	53
73	4	18, 18, 18, 18	11*, 13, 15, 15
75	1	40	30
77	1	60	44
79	1	78	55
81	1	54	40
83	1	82	62
85	4	16, 16, 16, 16	10*, 12, 12, 14
87	1	56	42
89	4	22, 22, 22, 22	15*, 15, 17, 19
91	3	24, 24, 24	16, 18*, 20
93	3	20, 20, 20	14*, 14, 16
95	1	72	54
97	2	48, 48	35*, 37
99	2	30, 30	22, 24*

REFERENCES

1. John Brillhart, D. H. Lehmer, J. L. Selfridge, Bryant Tuckerman, and S. S. Wagstaff, Jr., Factorizations of $b^n \pm 1$, $b = 2, 3, 5, 6, 7, 10, 11, 12$ up to high powers, American Mathematical Society, Contemporary Mathematics, Vol. 22, 1983.
2. Peter Hilton and Jean Pedersen, Approximating any regular polygon by folding paper, *Mathematics Magazine*, 56 (1983) 141–155.
3. Peter Hilton and Jean Pedersen, Folding regular star polygons and number theory, *The Mathematical Intelligencer*, 7 (1985) 15–26.
4. Peter Hilton and Jean Pedersen, On certain algorithms in the practice of geometry and the theory of numbers, *Publicacions Secció de Matemàtiques, Universitat Autònoma de Barcelona*, 29 (1985) 31–64.
5. Peter Hilton and Jean Pedersen, On generalized symbols, orders and quasi-orders, *Publicacions Secció de Matemàtiques, Universitat Autònoma de Barcelona*, 29 (1985) 123–144.
6. Peter Hilton and Jean Pedersen, The general quasi-order algorithm in number theory, *Internat. J. Math. and Math. Sci.*, 9 (1986) 245–251.
7. Peter Hilton and Jean Pedersen, On the complementary factor in a new congruence algorithm, *Internat. J. Math. and Math. Sci.*, to appear.
8. K. E. Kloss, Some number-theoretic calculations, *Journal of Research of the National Bureau of Standards*, 69B (1965) 335–336.

9. Jeffrey C. Lagarias, The $3x + 1$ problem and its generalizations, this MONTHLY 92 (1985) 3–23.
10. H. P. Lawther, Jr., An application of number theory to the splicing of telephone cables, this MONTHLY 42 (1935) 81–91.
11. Emma Lehmer, Generalizations of Gauss's Lemma, in *Number Theory and Algebra*, Academic Press, New York, 1977, 187–194.
12. D. P. Parent, Exercises in Number Theory, Springer-Verlag, New York, 1984.
13. J. B. Roberts, Integral power residues as permutations, this MONTHLY 76 (1969) 379–385.
14. Harold N. Shapiro, Introduction to the Theory of Numbers, Wiley, New York, 1983.
15. Wacław Sierpiński, Elementary Theory of Numbers (translated from Polish by A. Hulanicki), Państwowe Wydawnictwo Naukowe, Warsaw, 1964.
16. A. Wieferich, Zum letzten Fermat'schen Theorem, *J. reine u. angew. Math.*, 136 (1909), 293–302.
17. Man Keung Siu, When is -1 a power of 2?, *Mathematics Magazine*, 22 (1975) 284–286.

Integrals, an Introduction to Analytic Number Theory

ILAN VARDI, *Stanford University*

ILAN VARDI: I got my Ph.D. in Number Theory from M.I.T. in 1982, as a student of Dorian Goldfeld. I then spent a year at the Institute for Advanced Study. I was an acting assistant professor at Stanford from 1983 to 1985. After realizing that not everybody cared about Kloosterman Sums, I learned how to use a computer and tried out some applied math. I'm now interested in special functions related to determinants of Laplacians.

1. Introduction. An examination of Gradzhteyn and Ryzhyk's book of integral tables reveals a large number of difficult and obscure integral formulas. In my opinion one of the most remarkable is given on p. 532

$$\int_{\pi/4}^{\pi/2} \log\log\tan x\,dx = \frac{\pi}{2}\log\left[\frac{\Gamma\left(\frac{3}{4}\right)}{\Gamma\left(\frac{1}{4}\right)}\sqrt{2\pi}\right], \tag{1}$$

where

$$\Gamma(s) = \int_0^\infty e^{-t}t^{s-1}\,dt, \qquad s > 0$$

is the classical Γ-function. The reference given is to Bierens de Haan [2]. Failing to locate the proof of this formula, I decided to study equation (1) in some depth. It turns out that this formula requires some fairly involved analysis to prove, and also serves as a good example of how nontrivial number theory can be embedded in an integral formula.

The key to equation (1) is the *Dirichlet L-function*

$$L(s) = 1 - \frac{1}{3^s} + \frac{1}{5^s} - \frac{1}{7^s}\cdots.$$

This is a well-known function; for example every calculus student knows the formula

$$L(1) = 1 - \frac{1}{3} + \frac{1}{5}\cdots = \frac{\pi}{4}.$$

Also, by the alternating series test $L(s)$ converges for $0 < s < 1$. However, much more is known and Hurwitz proved that $L(s)$ can be analytically continued to an entire function in the whole complex plane. He did this by proving the *functional equation*

$$L(1-s) = \left(\frac{2}{\pi}\right)^s \sin\frac{\pi s}{2}\Gamma(s)L(s). \tag{2}$$

What we will, in fact, show is that

$$\int_{\pi/4}^{\pi/2} \log\log\tan x\,dx = \frac{d}{ds}\Gamma(s)L(s)\Big|_{s=1} \tag{3}$$

Invoking the well-known formulas

$$\Gamma(1) = 1$$
$$\Gamma'(1) = -\gamma,$$

where γ is *Euler's constant*,

$$\gamma = \lim_{n\to\infty}\left\{\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right) - \log n\right\} = .5772156649015328606065 12\ldots,$$

equation (3) becomes

$$\int_{\pi/4}^{\pi/2} \log\log\tan x\,dx = -\gamma\frac{\pi}{4} + L'(1).$$

So the proof of equation (1) will consist of 2 parts: a) establishing (3) b) expressing $L'(1)$ in terms of logarithms of Γ-functions.

2. Proof of equation (3). We begin with a general *Dirichlet series*

$$F(s) = \sum_{n=1}^{\infty}\frac{f(n)}{n^s},$$

which, if f is of polynomial growth, will converge absolutely in a half-plane $Re(s) > c$. We now use the technique first developed by Riemann to study the Riemann ζ-function

$$\Gamma(s) = \int_0^\infty e^{-t}t^{s-1}\,dt = \int_0^\infty e^{-nt}(nt)^{s-1}\,d(nt),$$

so

$$\frac{\Gamma(s)}{n^s} = \int_0^\infty e^{-nt}t^{s-1}\,dt.$$

Hence, by absolute convergence, one gets that for $Re(s) > c$

$$\Gamma(s)F(s) = \Gamma(s)\sum_{n=1}^{\infty}\frac{f(n)}{n^s} = \sum_{n=1}^{\infty} f(n)\int_0^\infty e^{-nt}t^{s-1}\,dt$$
$$= \int_0^\infty\left(\sum_{n=1}^{\infty} f(n)e^{-nt}\right)t^{s-1}\,dt.$$

Now let $z = e^{-t}$, this gives

$$\Gamma(s)F(s) = \int_0^1\left(\sum_{n=1}^{\infty} f(n)z^n\right)\left(\log\frac{1}{z}\right)^{s-1}\frac{dz}{z}.$$

Now we add the restriction that $f(n)$ be a *periodic* function. That is, there is a positive integer q such that $f(n+q)=f(n)$ for all n (for technical reasons also assume that $f(q)=0$). With these assumptions we have that for $|z|<1$

$$\sum_{n=1}^{\infty} f(n)z^n = \sum_{m=0}^{\infty}\sum_{n=1}^{q-1} f(mq+n)z^{mq+n}$$
$$= \frac{\sum_{n=1}^{q-1} f(n)z^n}{1-z^q} = \frac{P(z,f)}{1-z^q},$$

where

$$P(z,f) = \sum_{n=1}^{q-1} f(n)z^n.$$

We have thus obtained the formula:

$$F(s)\Gamma(s) = \int_0^1 \frac{P(z,f)\left(\log\frac{1}{z}\right)^{s-1}}{1-z^q}\,\frac{dz}{z}. \tag{4}$$

This formula was first obtained by Dirichlet (see [3]) to derive his *class number formula* of which $L(1)=\pi/4$ is the simplest case. Differentiating equation (4) by Leibniz's rule gives

$$\frac{d}{ds}F(s)\Gamma(s) = \int_0^1 P(z,f)\frac{\left(\log\frac{1}{z}\right)^{s-1}}{1-z^q}\log\log\left(\frac{1}{z}\right)\frac{dz}{z}.$$

Now if $F(s)$ converges absolutely at $s=1$ this will yield

$$F'(1)-\gamma F(1) = \int_0^1 P(z,f)\log\log\left(\frac{1}{z}\right)\frac{dz}{z}. \tag{5}$$

To prove equation (1) we let $q=4$ and pick $f(n)$ to be the *quadratic character* (mod 4) that is

$$\chi_4(n) = \begin{cases} 0 & \text{if } n\equiv 0 \pmod 4\\ 1 & \text{if } n\equiv 1 \pmod 4\\ 0 & \text{if } n\equiv 2 \pmod 4\\ -1 & \text{if } n\equiv 3 \pmod 4\end{cases}.$$

χ_4 is called the quadratic character (mod 4) because for $(n,4)=1$

$$\chi_4(n) = \begin{cases} 1 & \text{if } \exists x \text{ s.t. } x^2\equiv n \pmod 4\\ -1 & \text{otherwise,}\end{cases}$$

while $\chi_4(n)=0$ if $(n,q)>1$.

So we have

$$P(z, \chi) = z - z^3$$

and equation (5) becomes

$$L'(1) - \gamma\frac{\pi}{4} = \int_0^1 \frac{(z - z^3)\log\log\frac{1}{z}}{1 - z^4}\frac{dz}{z}$$

$$= \int_0^1 \log\log\left(\frac{1}{z}\right)\frac{dz}{1 + z^2} = \int_1^\infty \log\log u \frac{du}{1 + u^2}$$

$$= \int_{\pi/4}^{\pi/2} \log\log\tan x \, dx.$$

3. Evaluating $L'(1)$. It turns out that it is much easier, first, to evaluate $L'(0)$, then use the functional equation $L(s) \to L(1 - s)$ to obtain the value for $L'(1)$.

To compute $L'(0)$ we follow a method due to André Weil [7]. Let

$$\zeta(s, a) = \sum_{n=1}^{\infty} \frac{1}{(n + a)^s}, \qquad 0 < a \le 1,$$

be the *Hurwitz ζ-function*. It is easily shown to converge for $Re(s) > 1$. Using the integral formula [**8**]

$$\Gamma(s)\zeta(a, s) = \int_0^\infty \frac{e^{-at}}{1 - e^t} t^{s-1}\, dt$$

one can show that $\zeta(a, s)$ can be analytically continued to the whole complex plane with only a simple pole at $s = 1$. The relevance of $\zeta(a, s)$ is due to the formula

$$L(s) = 4^{-s}\left[\zeta\left(s, \frac{1}{4}\right) - \zeta\left(s, \frac{3}{4}\right)\right];$$

thus evaluating $\zeta'(0, a)$ will yield the value of $L'(0)$ (for ease of notation we have written $\zeta'(s, a)$ to denote $\frac{\partial}{\partial s}\zeta(s, a)$). Weil's observation is the following: note that for $s > 1$

$$\zeta(s, a + 1) = \zeta(s, a) - \frac{1}{a^s},$$

thus

$$\zeta'(s, a + 1) = \zeta(s, a) + a^{-s}\log a,$$

and at $s = 0$

$$\zeta(0, a + 1) = \zeta(0, a) + \log a.$$

Letting

$$G(a) = e^{\zeta'(0, a)},$$

we see that $G(a)$ satisfies the functional equation

$$G(a+1) = aG(a).$$

Further, one has that

$$\frac{d^2}{da^2}\log G(a) = \sum_{n=1}^{\infty} \frac{1}{(n+a)^2} > 0, \quad \text{for } a > 0,$$

and that

$$G(a) \text{ is analytic for } a > 0.$$

These however are the exact conditions for the *Bohr-Mollerup Theorem* for the uniqueness of the Gamma function **[1]**. Thus one has that

$$G(a) = G(1)\Gamma(a).$$

One sees that $G(1) = \zeta'(0,1)$, and on noting that $\zeta(s,1) = \zeta(s)$, where $\zeta(s)$ is the Riemann ζ-function, one has

$$G(1) = \zeta'(0).$$

It is well known that $\zeta'(0) = -(1/2)\log 2\pi$ (e.g., **[6]**, **[8]**), and so

$$\zeta'(0,a) = \log\frac{\Gamma(a)}{\sqrt{2\pi}}.$$

Substituting this in the formula for $L(s)$ one derives

$$L'(0) = \log\frac{\Gamma\left(\frac{1}{4}\right)}{\Gamma\left(\frac{3}{4}\right)} - L(0)\log 4.$$

By the functional equation and $L(1) = \pi/4$ one gets that

$$L(0) = \frac{1}{2}.$$

And once again by the functional equation

$$\frac{2}{\pi}\frac{d}{ds}\Gamma(s)L(s)\bigg|_{s=1} = \frac{1}{2}\log 4 + \log\frac{\Gamma\left(\frac{3}{4}\right)}{\Gamma\left(\frac{1}{4}\right)} + \frac{1}{2}\log\frac{\pi}{2},$$

and thus

$$L'(1) = \gamma\frac{\pi}{2} + \frac{\pi}{2}\log\frac{\Gamma\left(\frac{3}{4}\right)}{\Gamma\left(\frac{1}{4}\right)}\sqrt{2\pi}.$$

This concludes the proof of equation (1).

4. More formulas! There are actually quite a number of identities in Gradzhteyn and Ryzhyk similar to (1). For example, there are

$$\int_0^1 \log\log\left(\frac{1}{x}\right)\frac{dx}{1+x+x^2} = \frac{\pi}{\sqrt{3}}\log\left[\frac{\Gamma\left(\frac{2}{3}\right)}{\Gamma\left(\frac{1}{3}\right)}(2\pi)^{1/3}\right], \quad \text{page 571} \tag{6}$$

$$\int_0^1 \log\log\left(\frac{1}{x}\right)\frac{dx}{1-x+x^2} = \frac{2\pi}{\sqrt{3}}\left[\frac{5}{6}\log 2\pi - \log\Gamma\left(\frac{1}{6}\right)\right], \quad \text{page 572.} \tag{7}$$

One sees that in equation (6) 3 plays the "key role" and in equation (7) 6 is the "magic number." To explain this one introduces *Dirichlet characters* (mod q)

$$\begin{aligned} &\chi \text{ is a Dirichlet character (mod } q) \text{ if} \\ \chi(1) &= 1 \\ \chi(n+q) &= \chi(n) && \forall n \\ \chi(n) &= 0 && \text{if } (n,q) > 1 \\ \chi(mn) &= \chi(m)\chi(n) && \forall m, n. \end{aligned}$$

The corresponding Dirichlet L-function is

$$L(s,\chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}, \quad Re(s) > 0$$

and can be continued to an entire function if χ is not the *trivial character*

$$\chi_0(n) = 1 \quad \text{if } (n,q) = 1$$

Now the analogous character to χ_4 in equation (6) is the quadratic character (mod 3)

$$\chi_3(n) = \begin{cases} 0 & \text{if } n \equiv 0 \pmod 3 \\ 1 & \text{if } n \equiv 1 \pmod 3 \\ -1 & \text{if } n \equiv 2 \pmod 3, \end{cases}$$

and in equation (7) the corresponding character is the quadratic character (mod 6) $\chi_6(n)$. Hence we have the L-functions

$$L(s,\chi_3) = 1 - \frac{1}{2^s} + \frac{1}{4^s} - \frac{1}{5^s} \cdots$$

$$L(s,\chi_6) = 1 - \frac{1}{5^s} + \frac{1}{7^s} \cdots.$$

The proofs of (6) and (7) are completely analogous to our proof of equation (1).

One can further explain how the numbers 3, 4, 6 play the key roles in our formulas. First rewrite equation (1) in the same form as (6) and (7)

$$\int_0^1 \log\log\left(\frac{1}{x}\right)\frac{dx}{1+x^2} = \frac{\pi}{2}\log\left[\frac{\Gamma\left(\frac{3}{4}\right)}{\Gamma\left(\frac{1}{4}\right)}\sqrt{2\pi}\right].$$

Note that the solutions of x^2+1 are 4th roots of unity, i and $-i$, and one explains why $L(s,\chi_4)$ is involved by noting that it can be shown from the *Quadratic Reciprocity Theorem* that

$$\zeta_{\mathbf{Q}(i)}(s) = L(s,\chi_4)\zeta(s),$$

where $\zeta_{\mathbf{Q}(i)}(s)$ is the *Dedekind zeta function* of the field $\mathbf{Q}(i)$, and the classical definition (e.g., **[5]**) of the Dedekind ζ-function of the number field K is

$$\zeta_K(s) = \sum_{\substack{A\subseteq K\\ A \text{ ideal}}} \frac{1}{N(A)^s}.$$

Similarly, x^2+x+1 is the irreducible polynomial for the 3rd roots of unity, $-1/2 \pm \dfrac{\sqrt{-3}}{2}$, and, as above, $L(s,\chi_3)$ appears because

$$\zeta_{\mathbf{Q}(\sqrt{-3})}(s) = L(s,\chi_3)\zeta(s).$$

Similarly, x^2-x+1 gives the 6th roots of 1, so, as above, one expects $L(s,\chi_6)$ to play the central role.

5. Exercises.

1) Show that

$$\int_1^e \log(-\log\log y)\,dy = -\sum_{n=1}^{\infty}\frac{\log n}{n!} - \gamma e.$$

Hint: consider

$$L_f(s) = \sum_{n=1}^{\infty}\frac{1}{(n-1)!n^s}.$$

2) Find a similar formula for

$$\int_e^{e^e} \log(-\log\log\log y)\,dy.$$

REFERENCES

1. E. Artin, The Gamma Function, Holt, Rinehart and Winston, New York, 1964.
2. D. Bierens de Haan, Nouvelles Tables d'intégrales définies, Amsterdam, 1867.
3. H. Davenport, Multiplicative Number Theory, Springer Verlag, New York, 1980.
4. I. S. Gradzhteyn and I. M. Ryzhyk, Table of integrals, Series, and Products, Academic Press, New York, 1980.
5. E. Hecke, Lectures on the Theory of Algebraic Numbers, Springer Verlag, New York, 1981.
6. I. Vardi, Determinants of Laplacians and Multiple Gamma Functions, preprint, 1986.
7. A. Weil, Elliptic Functions According to Eisenstein and Kronecker, Springer Verlag, New York, 1976.
8. E. T. Whittaker and G. N. Watson, A Course of Modern Analysis, Cambridge University Press, 1965.

Dice with Fair Sums

LEWIS C. ROBERTSON: B.S., M.S. (University of Chicago), Ph.D. (UCLA).

RAE MICHAEL SHORTT: I received a B.A. in mathematics at Amherst College in 1978 and my doctorate at M.I.T. in 1982.
My interests include measure theory, probability, and descriptive set theory.

STEPHEN G. LANDRY: I received my B.A. in mathematics from the University of New Haven in 1984. Since then I have been a graduate student at Wesleyan University.

Consider the following situation. We are given two n-sided dice with the numbers 1 through n on their sides. These dice have the property that when both are rolled independently, the sum of the numbers showing behaves (as a random variable) as if the two dice were each fair. Can we conclude that each die is in fact fair? It surprised the authors to discover that the answer is sometimes no, depending on the particular value of n being considered. What follows is an exposition of this discovery.

Let X be a random variable. Suppose that X takes on only finitely many values and that each of these values is a positive integer. Let n be the largest integer such that $P(X = n) > 0$. We then call X an *n-sided die*. Define the polynomial $p_X(x) = r_0 + r_1x + \cdots + r_{n-1}x^{n-1}$, where

$$r_k = \frac{P(X = k+1)}{P(X = n)} \quad \text{for } k = 0, 1, \ldots, n-1.$$

We call p_X the *companion polynomial* for the random variable X. It is easy to see that a polynomial is the companion polynomial for some die if and only if it is monic and has nonnegative coefficients. Then

$$P(X = k+1) = \frac{r_k}{r_0 + r_1 + \cdots + r_{n-1}} \quad \text{for } k = 0, 1, \ldots, n-1.$$

What we have called a companion polynomial is simply a normalized version of the generating function (see [**1**], Chapter 11). An n-sided die is *fair* if its companion is

the *cyclic polynomial*

$$f_n(x) = 1 + x \cdots + x^{n-1}.$$

LEMMA 1.1. *Let X and Y be, respectively, n- and m-sided dice with companion polynomials p_X and p_Y. If X and Y are independent, then $W = X + Y$ is an $(n + m)$-sided die whose companion polynomial is $p_W(x) = xp_X(x)p_Y(x)$.*

Proof. We put

$$\begin{aligned} p_X(x) &= r_0 + r_1 x + \cdots + r_{n-1}x^{n-1} \\ p_Y(x) &= s_0 + s_1 x + \cdots + s_{m-1}x^{m-1} \\ p_X(x)p_Y(x) &= t_0 + t_1 x + \cdots + t_{n+m-2}x^{n+m-2} \\ p_W(x) &= u_0 + u_1 x + \cdots + u_{n+m-1}x^{n+m-1}. \end{aligned}$$

Then for $k = 0, 1, \ldots, n + m - 1$,

$$\begin{aligned} u_k = \frac{P(W = k+1)}{P(W = n+m)} &= \sum_{i=1}^{n} \frac{P(X = i \text{ and } Y = k + 1 - i)}{P(W = n + m)} \\ &= \sum_{i=1}^{n} \frac{P(X = i)}{P(X = n)} \cdot \frac{P(Y = k + 1 - i)}{P(Y = m)} \\ &= \sum_{i=1}^{n} r_{i-1}s_{k-i} = \sum_{i=0}^{n-1} r_i s_{k-1-i}. \end{aligned}$$

Note that because n and m are the greatest possible values of X and Y, respectively, it follows that

$$P(W = n + m) = P(X = n \text{ and } Y = m) = P(X = n)P(Y = m).$$

Above, we have examined all the ways in which $X + Y$ can equal $k + 1$, and then used the independence of X and Y. Multiplication of the polynomials p_X and p_Y yields

$$t_{k-1} = \sum_{i=0}^{n-1} r_i S_{k-1-i},$$

where we interpret the $r_i = 0$ for $i > n$ or $i < 1$ and $s_i = 0$ for $i > m$ or $i < 1$. The equality $t_{k-1} = u_k$ implies that

$$p_W(x) = xp_X(x)p_Y(x).$$

Call a polynomial $p(x)$ *palindromic* if its coefficients are the same when read forwards and backwards. Equivalently, we may define a polynomial p of degree n to be palindromic if $p(x) = x^n p(x^{-1})$. It is not hard to see that the product of two palindromic polynomials is again palindromic.

If two random variables X and Y have the same distribution (or "law"), we write $L(X) = L(Y)$. Note that two dice have the same distribution if and only if they have the same companion polynomial.

The following result answers one half of our questions about dice whose sum is fair.

THEOREM 1.2. *Suppose that* $n \in \{1,2,\ldots,9\} \cup \{11,13\}$. *Let* X *and* Y *be independent fair n-sided dice. Let* U *and* V *be independent dice, each of which has* $\leqslant n$ *sides. Suppose that* $L(U+V) = L(X+Y)$. *Then* U *and* V *are fair n-sided dice.*

Proof. This involves a certain amount of case-checking. For example, suppose $n = 6$. Let p_X, p_Y, p_U, p_V be the companion polynomials for X, Y, U, V, respectively. The companion polynomial for $X + Y$ and $U + V$ is (by lemma 1.1)

$$xp_U(x)p_V(x) = xp_X(x)p_Y(x) = x[f_6(x)]^2.$$

Now $f_6(x) = (1+x)q_1(x)q_2(x)$, where $q_1(x) = 1 + x + x^2$ and $q_2(x) = 1 - x + x^2$ are irreducible quadratic polynomials. Considering the degrees of these polynomials, we see that $deg(p_U) = deg(p_V) = 5$. Only two combinations of the irreducible factors $(1+x)$, $q_1(x)$, $q_2(x)$ are possible:

$$p_U(x) = p_V(x) = (1+x)q_1(x)q_2(x) = f_6(x)$$

or

$$p_U(x) = (1+x)(q_1(x))^2 \qquad p_V(x) = (1+x)(q_2(x))^2.$$

However, $(1+x)(q_2(x))^2$ has a couple of negative coefficients and hence cannot be the companion polynomial of any random variable. We are left with the case $p_U = p_V = p_X = p_Y$. Then U and V are fair 6-sided dice.

A similar procedure must be carried out for each n. One writes

$$f_n(x) = \begin{cases} (1+x)q_1(x)\cdots q_m(x) & n \text{ even} \\ q_1(x)\cdots q_m(x) & n \text{ odd} \end{cases},$$

where

$$m = \begin{cases} \dfrac{n-2}{2} & n \text{ even} \\ \dfrac{n-1}{2} & n \text{ odd} \end{cases},$$

and $q_1 \cdots q_m$ are irreducible quadratic factors. One has $q_p(x) = 1 - k_n(p)x + x^2$, where $k_n(p) = 2\cos(2p\pi/n)$ for $p = 1,\ldots,m$. All of the real factorizations of $[f_n(x)]^2$ into products of these irreducible factors must be calculated.

For each n, these cases were examined (for the presence of negative coefficients) using a computer and MACSYMA, a large symbolic manipulation program developed at the M.I.T. Laboratory for Computer Science. In each case, it was determined that only the "fair" factorizations of $[f_n(x)]^2$ yielded polynomials with nonnegative coefficients. A copy of the computer output is available upon request.

The theorem is no longer true if we relax the condition that U and V have $\leqslant n$ sides. For example, suppose $n = 6$. Allowing the dice U and V to have more than

six spots on a given side, we may consider a factorization

$$[f_6(x)]^2 = p_U(x)p_Y(x) = [(1+x)q_1(x)] \cdot \left[(1+x)q_1(x)(q_2(x))^2\right],$$
$$= [1 + 2x + 2x^2 + x^3][1 + x^2 + x^3 + x^4 + x^5 + x^7],$$

which yields the following.

Example 1.3. Let U and V be independent random variables with distributions given by

Event	$U=1$	$U=2$	$U=3$	$U=4$
Probability	1/6	2/6	2/6	1/6

Event	$V=1$	$V=2$	$V=3$	$V=4$	$V=5$	$V=6$	$V=7$	$V=8$
Probability	1/6	0	1/6	1/6	1/6	1/6	0	1/6

Then $W = U + V$ behaves as if it were the sum of two fair dice.

Event	$W=2$	$W=3$	$W=4$	$W=5$	$W=6$	$W=7$	$W=8$	$W=9$	$W=10$	$W=11$	$W=12$
Probability	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

W. W. Funkenbusch pointed out to the authors that the dice U and V can be physically realized by taking a pair of ordinary dice and changing the number of spots on their sides.

U:	1	2	2	3	3	4

V:	1	3	4	5	6	8

Of course, according to our definition of "n-sided die," given above, the dice U and V here are "4-sided" and "8-sided," respectively.

THEOREM 1.4. *Let X, Y, U, V be n-sided dice. Suppose that X and Y are independent fair dice and that $L(U + V) = L(X + Y)$. Let p_U and p_V be the companion polynomials for U and V. Then*

1) *p_U and p_V are palindromic.*
2) *If $p_U = p_V$, then U and V are fair dice.*
3) *Suppose that the coefficients of p_U and p_V are all positive rational numbers. Then U and V are fair dice.*

Proof. 1) It follows from lemma 1.1 that $p_U(x)p_V(x) = [f_n(x)]^2$. Now the complex roots of f_n are the $n - 1$ roots of unity, excluding $x = 1$. We pair each root with its complex conjugate and write

$$f_n(x) = \begin{cases} (x+1)q_1(x) \cdots q_m(x) & n \text{ even} \\ q_1(x) \cdots q_m(x) & n \text{ odd} \end{cases}$$

where $q_1, \ldots, q_m$ are irreducible quadratics of the form $q(x) = (x - \omega)(x - \omega) = 1 + kx + x^2$. The irreducible factors of p_U and p_V are thus all palindromic. So p_U and p_V are palindromic.

2) In case $p_U = p_V$, we see that $p_U(x)p_V(x) = [p_U(x)]^2 = [f_n(x)]^2$, so that $(p_U - f_n)(p_U + f_n)$ is the zero polynomial. It follows that $p_U = p_V = f_n$, so that U and V are fair dice.

3) The roots of $p_U(x)$ and $p_V(x)$ are algebraic integers. Gauss' lemma [**3**; Theorem 9.7] implies that the coefficients of p_U and p_V are ordinary ("rational") integers. Put

$$p_U(x) = r_0 + r_1x + r_2x^2 + \cdots + r_{n-2}x^{n-2} + r_{n-1}x^{n-1}$$

$$p_V(x) = s_0 + s_1x + s_2x^2 + \cdots + s_{n-2}x^{n-2} + s_{n-1}x^{n-1}$$

$$[f_n(x)]^2 = t_0 + t_1x + t_2x^2 + \cdots + t_{2n-3}x^{2n-3} + t_{2n-2}x^{2n-2}.$$

Then $t_k = \min\{k+1, 2n-1-k\} = r_0s_k + r_1s_{k-1} + \cdots + r_ks_0$. Now p_U and p_V are monic, and from part 1, palindromic. So $r_{n-1} = 1 = r_0$ and $s_{n-1} = 1 = s_0$. Then $t_1 = 2 = r_0s_1 + r_1s_0 = s_1 + r_1$. Since all the coefficients are assumed to be positive, the only possibility is that $1 = s_1 = s_{n-2}$ and $1 = r_1 = r_{n-2}$. Then consider $t_2 = 3 = r_0s_2 + r_1s_1 + r_2s_0 = s_2 + 1 + r_2$, forcing $s_2 = s_{n-3} = 1$ and $r_2 = r_{n-3} = 1$. One proceeds inductively to obtain $r_0 = r_1 = \cdots = r_{n-1} = s_0 = s_1 = \cdots = s_{n-1} = 1$. Both U and V are fair dice.

In part 3 of theorem 1.4, the assumption of rationality for the coefficients cannot be eliminated. In theorem 1.14 below, we exhibit independent, but unfair, n-sided dice whose sum behaves as if the dice were fair.

One further note: the coefficients $r_0, r_1, \ldots, r_{n-1}$ are all rational if and only if $s_0, s_1, \ldots, s_{n-1}$ are all rational. This is because, when $r_0, r_1, \ldots, r_{n-1}$ are known, the coefficients $s_0, s_1, \ldots, s_{n-1}$ may be obtained by solving the system of linear equations

$$r_0s_k + r_1s_{k-1} + \cdots + r_ks_0 = \min\{k+1, 2n-1-k\}$$

using Cramer's Rule. Since the matrix of coefficients and the $\min\{k+1, 2n-1-k\}$ are rational, it follows that $s_0, \ldots, s_{n-1}$ are rational.

We now turn our attention to Chebyshev polynomials. They are a most useful computational tool for handling the factorization of cyclic polynomials.

The Chebyshev polynomials $S_0, S_1, S_2, \ldots$ (actually, these are versions of what are termed "Chebyshev polynomials of the first kind") have many uses in mathematical analysis, and there are many equivalent methods of defining them (see, e.g. [**2**] or [**4**]). One way is to set $S_0(x) = 1$ and $S_1(x) = x$ and use the recurrence relation $S_{j+1}(x) = xS_j(x) - S_{j-1}(x)$. A matrix formulation is

$$\begin{bmatrix} 0 & -1 \\ 1 & x \end{bmatrix}^n = \begin{bmatrix} -S_{n-2}(x) & -S_{n-1}(x) \\ S_{n-1}(x) & S_n(x) \end{bmatrix}.$$

We summarize the basic facts about Chebyshev polynomials that we shall need in

LEMMA 1.5. *Let $S_0, S_1, S_2, \ldots$ be the sequence of Chebyshev polynomials.*

1) $(x-2)\sum_{m=0}^{j-1} S_m(x) = S_j(x) - S_{j-1}(x) - 1$ *for* $j = 1, 2, \ldots$.

2) $(\sin\theta)S_{j-1}(2\cos\theta) = \sin(j\theta)$ *for* $j = 1, 2, \ldots$.

3) *Let α and β be real numbers in $[-1,1]$ such that neither $1-\alpha z+z^2$ nor $1-\beta z+z^2$ vanishes for complex values z in the open disk $|z|<1/2$. Then for $|z|<1/2$*

$$\frac{1-\alpha z+z^2}{1-\beta z+z^2}=1+(\beta-\alpha)\sum_{j=0}^{\infty}S_j(\beta)z^{j+1}.$$

Proof. 1) Define the matrix

$$A=\begin{bmatrix}0 & -1\\ 1 & x\end{bmatrix},$$

and consider the identity

$$\begin{aligned}I+A+A^2+\cdots+A^{n-1}&=(A-I)^{-1}(A^n-I)\\ &=\frac{1}{x-2}\begin{bmatrix}1-x & -1\\ 1 & 1\end{bmatrix}\begin{bmatrix}-S_{n-2}(x)-1 & -S_{n-1}(x)\\ S_{n-1}(x) & S_n(x)-1\end{bmatrix}.\end{aligned}$$

Inspection of the lower right-hand entry yields the desired result.

2) This well-known identity follows from the basic recurrence relation by a simple induction argument.

3) Define a function of two variables

$$g(z,\beta)=\sum_{j=0}^{\infty}S_j(\beta)z^j.$$

Then

$$\begin{aligned}(\beta-z)g(z,\beta)&=\beta+\sum_{j=1}^{\infty}\left(\beta S_j(\beta)-S_{j-1}(\beta)\right)z^j\\ &=\beta+\sum_{j=1}^{\infty}S_{j+1}(\beta)z^j\end{aligned}$$

by the recurrence relation. So

$$z(\beta-z)g(z,\beta)-\beta z=\sum_{j=1}^{\infty}S_{j+1}(\beta)z^{j+1}=g(z,\beta)-1-\beta z,$$

which leads to

$$g(z,\beta)=\sum_{j=0}^{\infty}S_j(\beta)z^j=\frac{1}{z^2-\beta z+1}.$$

Now

$$\frac{1-\alpha z+z^2}{1-\beta z+z^2}-1=\frac{(\beta-\alpha)z}{1-\beta z+z^2}=(\beta-\alpha)zg(z,\beta),$$

so that

$$\frac{1-\alpha z+z^2}{1-\beta z+z^2}=1+(\beta-\alpha)\sum_{j=0}^{\infty}S_j(\beta)z^{j+1}$$

as desired. It remains only to note that $|S_n(\beta)|\leqslant 2^n$ whenever $|\beta|\leqslant 1$ (proof by induction), so that the power series we have been manipulating so freely in fact converges for $|z|<1/2$. (The reader might want to press on and show convergence for $|z|<1$, which does hold.)

In view of theorem 1.2, it involves no loss of generality in considering n-sided dice with fair sums where $n\geqslant 10$. With this restriction on n, we introduce some notation. Given $n\geqslant 10$, choose $\{p,q\}\subseteq\{1,2,\dots,m\}$, where $p\neq q$ and

$$m=\begin{cases}\dfrac{n-1}{2} & \text{if } n \text{ is odd}\\[2mm] \dfrac{n-2}{2} & \text{if } n \text{ is even.}\end{cases}$$

Define $k_n(p)=2\cos(2\pi p/n)$ and set

$$h_n(p,q;x)=\frac{1-k_n(p)x+x^2}{1-k_n(q)x+x^2}f_n(x).$$

Put

$$\begin{aligned}\lambda_n(p,q)&=\cos(2\pi q/n)-\cos(2\pi p/n)\\&=2\sin(\pi(p+q)/n)\sin(\pi(p-q)/n),\end{aligned}$$

and define $c_j=c_j(p,q,n)$ by

$$h_n(p,q;x)=c_0+c_1x+c_2x^2+\cdots+c_{n-1}x^{n-1}$$

and $d_j=d_j(p,q,n)=c_j(q,p,n)$, so that

$$h_n(q,p;x)=d_0+d_1x+d_2x^2+\cdots+d_{n-1}x^{n-1}.$$

Define

$$\phi_j(p)=\sin\left(\frac{2p\pi}{n}\right)+\sin\left(\frac{2jp\pi}{n}\right)-\sin\left(\frac{2(j+1)p\pi}{n}\right)$$

THEOREM 1.6. *With notation as above, the following identities hold*:

$$c_j=1+\lambda_n(p,q)\left(1-\cos\left(\frac{2q\pi}{n}\right)\right)^{-1}\phi_j(q)\left(\sin\left(\frac{2q\pi}{n}\right)\right)^{-l}$$

$$d_j=1+\lambda_n(q,p)\left(1-\cos\left(\frac{2p\pi}{n}\right)\right)^{-1}\phi_j(p)\left(\sin\left(\frac{2p\pi}{n}\right)\right)^{-1}$$

Proof. Since $1 - k_n(p)z + z^2$ and $1 - k_n(q)z + z^2$ are nonzero in the complex disk $|z| < 1/2$, we have

$$\frac{1 - \alpha x + x^2}{1 - \beta x + x^2} = 1 + (\beta - \alpha) \sum_{j=0}^{\infty} S_j(\beta) x^{j+1} \qquad |x| < 1/2$$

for $\alpha = k_n(p)$ and $\beta = k_n(q)$ or *vice versa* (lemma 1.5.3). So

$$\begin{aligned} h_n(p, q; x) &= \{1 + 2\lambda_n(p,q)(S_0(k_n(q))x + S_1(k_n(q))x^2 + \cdots)\} \\ &\quad \times \{1 + (x + x^2 + \cdots + x^{n-1})\} \\ &= \{1 + A(x)\} \times \{1 + B(x)\} \\ &= \{1 + A(x) + B(x) + A(x)B(x)\}. \end{aligned}$$

We calculate

$$\begin{aligned} A(x)B(x) &= 2\lambda_n(p,q)(S_0(k_n(q))x + S_1(k_n(q))x^2 + \cdots) \\ &\quad \times (x + x^2 + \cdots + x^{n-1}) \\ &= \lambda_n(p,q)(S_0(k_n(q))x^2 + (S_0(k_n(q)) + S_1(k_n(q)))x^3 + \cdots). \end{aligned}$$

Then

$$\begin{aligned} h_n(p,q;x) = 1 &+ x(2\lambda_n(p,q)S_0(k_n(q)) + 1) \\ &+ x^2(2\lambda_n(p,q)(S_0(k_n(q)) + S_1(k_n(q))) + 1) \\ &+ x^3(2\lambda_n(p,q)(S_0(k_n(q)) + S_1(k_n(q)) + S_2(k_n(q))) + 1) \\ &+ \cdots, \end{aligned}$$

so that

$$\begin{aligned} c_j &= 1 + 2\lambda_n(p,q) \sum_{m=0}^{j-1} S_m(k_n(q)) \\ &= 1 + 2\lambda_n(p,q)(k_n(q) - 2)^{-1}(S_j(k_n(q)) - S_{j-1}(k_n(q)) - 1) \end{aligned}$$

for $j = 0, 1, \ldots, n-1$ (lemma 1.5.1). Note that this infinite series expansion of $h_n(p, q; x)$ is in fact just a polynomial. Application of lemma 1.5.2 shows that

$$\begin{aligned} c_j &= 1 + 2\lambda_n(p,q)(k_n(q) - 2)^{-1}\left(\frac{\sin((j+1)2\pi q/n)}{\sin(2\pi q/n)} - \frac{\sin(2\pi qj/n)}{\sin(2\pi q/n)} - 1\right) \\ &= 1 + \lambda_n(p,q)\left(1 - \cos\left(\frac{2\pi q}{n}\right)\right)^{-1} \phi_j(q)(\sin(2\pi q/n))^{-1} \end{aligned}$$

as desired. The formula for d_j is obtained by reversing the roles of p and q.

THEOREM 1.7. *With notation as above, one has $c_j > 0$ if and only if*

$$\lambda_n(q,p)\left(\cos\left(\frac{q\pi}{n}\right) - \cos\left(\frac{(2j+1)q\pi}{n}\right)\right) < \sin\left(\frac{q\pi}{n}\right)\sin\left(\frac{2q\pi}{n}\right)$$

and $d_j > 0$ *if and only if*

$$\lambda_n(q, p)\left(\cos\left(\frac{(2j+1)p\pi}{n}\right) - \cos\left(\frac{p\pi}{n}\right)\right) < \sin\left(\frac{p\pi}{n}\right)\sin\left(\frac{2p\pi}{n}\right).$$

Proof. Using theorem 1.6, we note the following equivalent statements, each of which is equivalent with $c_j > 0$:

$$-\lambda_n(p, q)\phi_j(q) < \left(1 - \cos\left(\frac{2\pi q}{n}\right)\right)\sin\left(\frac{2\pi q}{n}\right)$$

$$-\lambda_n(p, q)\phi_j(q) < 2\left(\sin\left(\frac{q\pi}{n}\right)\right)^2\sin\left(\frac{2\pi q}{n}\right)$$

$$-\lambda_n(p, q)\left(2\sin\left(\frac{q\pi}{n}\right)\cos\left(\frac{q\pi}{n}\right) + 2\cos\left(\frac{(2j+1)q\pi}{n}\right)\sin\left(-\frac{q\pi}{n}\right)\right)$$

$$< 2\left(\sin\left(\frac{q\pi}{n}\right)\right)^2\sin\left(\frac{2\pi q}{n}\right)$$

$$\lambda_n(q, p)\left(\cos\left(\frac{q\pi}{n}\right) - \cos\left(\frac{(2j+1)q\pi}{n}\right)\right) < \sin\left(\frac{q\pi}{n}\right)\sin\left(\frac{2q\pi}{n}\right),$$

as desired. The result for d_j follows in like manner.

COROLLARY 1.8. *With notation as above, suppose that* $p < q$, *so that* $\lambda_n(p, q) < 0 < \lambda_n(q, p)$. *Then* $h_n(p, q; x)$ *has strictly positive coefficients if*

$$2\sin\left(\frac{(p+q)\pi}{n}\right)\sin\left(\frac{(q-p)\pi}{n}\right) < \tan\left(\frac{q\pi}{2n}\right)\sin\left(\frac{2q\pi}{n}\right), \tag{*}$$

and $h_n(q, p; x)$ *has strictly positive coefficients if*

$$2\sin\left(\frac{(p+q)\pi}{n}\right)\sin\left(\frac{(q-p)\pi}{n}\right)\tan\left(\frac{p\pi}{2n}\right) < \sin\left(\frac{2p\pi}{n}\right). \tag{**}$$

Proof. Assuming (*) holds, we have

$$\begin{aligned}\lambda_n(q, p)\left(\cos\left(\frac{q\pi}{n}\right) - \cos\left(\frac{(2j+1)q\pi}{n}\right)\right) &< \lambda_n(q, p)\left(\cos\left(\frac{q\pi}{n}\right) + 1\right)\\ &< \sin\left(\frac{2q\pi}{n}\right)\tan\left(\frac{q\pi}{2n}\right)\left(\cos\left(\frac{q\pi}{n}\right) + 1\right)\\ &= \sin\left(\frac{2q\pi}{n}\right)\sin\left(\frac{q\pi}{n}\right),\end{aligned}$$

so that, by theorem 1.7, each c_j is positive. Now assume (**) and note

$$\begin{aligned}\lambda_n(q,p)\left(\cos\left(\frac{(2j+1)p\pi}{n}\right)-\cos\left(\frac{p\pi}{n}\right)\right) &< \lambda_n(q,p)\left(1-\cos\left(\frac{p\pi}{n}\right)\right)\\ &< \lambda_n(q,p)\tan\left(\frac{p\pi}{2n}\right)\sin\left(\frac{p\pi}{n}\right)\\ &< \sin\left(\frac{2p\pi}{n}\right)\sin\left(\frac{p\pi}{n}\right),\end{aligned}$$

so that, by theorem 1.7, each d_j is positive.

Define a function Q by the rule

$$Q(a,b,n)=\frac{\sin(a\pi/n)}{2\sin(b\pi/n)}.$$

COROLLARY 1.9. *With notation as above, assume that $q = p + 1 < n/2$. Then $h_n(p, q; z)$ has strictly positive coefficients if*

$$\sin\left(\frac{\pi}{n}\right) < \tan\left(\frac{q\pi}{2n}\right)Q(2p+2, 2p+1, n),$$

and $h_n(q, p; x)$ has strictly positive coefficients if

$$\sin\left(\frac{\pi}{n}\right)\tan\left(\frac{p\pi}{2n}\right) < Q(2p, 2p+1, n).$$

Proof. Immediate consequence of corollary 1.8.

COROLLARY 1.10. *Given $n \geqslant 10$, take $p = (n-3)/2$, $q = (n-1)/2$ if n is odd; take $p = (n-4)/2$, $q = (n-2)/2$ if n is even. Define $\delta(n) \in \{1,2\}$ by $2p + \delta(n) = n - 2$. Then $h_n(p, q; x)$ has strictly positive coefficients if*

$$\sin\left(\frac{\pi}{n}\right) < \tan\left(\frac{q\pi}{n}\right)Q(\delta(n), \delta(n)+1, n).$$

Also $h_n(q, p; x)$ has strictly positive coefficients if

$$\sin\left(\frac{\pi}{n}\right)\tan\left(\frac{p\pi}{2n}\right) < Q(\delta(n)+2, \delta(n)+1, n).$$

Proof. This follows from corollary 1.9, noting that

$$\sin\left(\frac{(2p+2)\pi}{n}\right) = \sin\left(\frac{\delta(n)\pi}{n}\right)$$

$$\sin\left(\frac{(2p+1)\pi}{n}\right) = \sin\left(\frac{(\delta(n)+1)\pi}{n}\right)$$

$$\sin\left(\frac{2p\pi}{n}\right) = \sin\left(\frac{(\delta(n)+2)\pi}{n}\right).$$

THEOREM 1.11. *Given* $n \geqslant 10$, *put* $\delta(n) = 1$ *if* n *is odd and* $\delta(n) = 2$ *if* n *is even. Set* $m = (n - \delta(n))/2$. *Then* $h_n(m, m-1; x)$ *has strictly positive coefficients.*

Proof. By corollary 1.10, it suffices to show that

$$\sin\left(\frac{\pi}{n}\right)\tan\left(\frac{p\pi}{2n}\right) < Q(\delta(n)+2, \delta(n)+1, n).$$

Now $Q(\delta(n)+2, \delta(n)+1, n) > 1/2$ because

$$\sin\left(\frac{(\delta(n)+2)\pi}{n}\right) > \sin\left(\frac{(\delta(n)+1)\pi}{n}\right) > 0.$$

Also,

$$\tan\left(\frac{p\pi}{2n}\right) < \tan\left(\frac{\pi}{4}\right) = 1$$

because $2p \leqslant n - 3$. Thus

$$\sin\left(\frac{\pi}{n}\right)\tan\left(\frac{p\pi}{2n}\right) < \sin\left(\frac{\pi}{n}\right) < 1/2 < Q(\delta(n)+2, \delta(n)+1, n),$$

as desired.

LEMMA 1.12. *If* $n \geqslant 10$, *then* $Q(2, 3, n) > 1/3$.

Proof. For $\theta = \pi/n$, we need to show that $(\sin 2\theta)/(\sin 3\theta) > 2/3$. Now

$$\cos\theta + \sin\theta\left(\frac{\cos 2\theta}{\sin 2\theta}\right) < 3/2$$

because $(\cos 2\theta)/\cos\theta < 1$, and hence

$$\sin 3\theta = \sin 2\theta\cos\theta + \sin\theta\cos 2\theta < (3/2)\sin 2\theta.$$

The results follow, noting that $0 < \sin 2\theta < \sin 3\theta$.

THEOREM 1.13. *Given* $n \geqslant 10$, *put* $\delta(n) = 1$ *if* n *is odd and* $\delta(n) = 2$ *if* n *is even. Set* $m = (n - \delta(n))/2$. *Then* $h_n(m-1, m; x)$ *has strictly positive coefficients for* $n \geqslant 14$ *and for* $n = 12$.

Proof. By corollary 1.9, it suffices to show that

$$\sin\left(\frac{\pi}{n}\right) < \tan\left(\frac{q\pi}{2n}\right)Q(\delta(n), \delta(n)+1, n),$$

where $q = m$. For n odd, this is the same as $\sin(2\pi/n) < (1/2)\tan(m\pi/2n)$. For n even, lemma 1.12 serves to establish the sufficiency of $\sin(\pi/n) < (1/3)\tan(m\pi/2n)$. In either case, success at a given value n_0 implies success at all $n \geqslant n_0$ such that $n \equiv n_0 \pmod 2$. Thus the proof is complete for $n \geqslant 14$ upon inspection of the

following table of calculator approximations:

n	$\sin(\pi/n)$	$\sin(2\pi/n)$	$(1/3)\tan(m\pi/2n)$	$(1/2)\tan(m\pi/2n)$
12	.2588		.2558	
13		.4647		.4430
14	.2225		.2658	
15		.4067		.4502

For $n = 12$, we note

$$\sin\left(\frac{\pi}{12}\right) < .26 < .27 < \tan\left(\frac{5\pi}{24}\right)Q(2,3,12).$$

Finally, we are ready for a result complementary to theorem 1.2. Combined, the two results provide a complete answer to the question of whether unfair n-sided dice can have a fair sum.

THEOREM 1.14. *Suppose that $n \in \{10, 12\} \cup \{14, 15, \dots\}$. Let X and Y be independent fair n-sided dice. There are independent n-sided dice U and V such that $L(U+V) = L(X+Y)$, but U and V are not fair.*

Proof. **Case 1**: Suppose $n \in \{12\} \cup \{14, 15, \dots\}$. Put $\delta = 1$ if n is odd and $\delta = 2$ if n is even. Set $m = (n - \delta)/2$. By theorems 1.11 and 1.13, $h_n(m-1, m; x)$ and $h_n(m, m-1; x)$ are polynomials with strictly positive coefficients. So let U and V be n-sided dice whose companion polynomials are

$$p_U(x) = h_n(m-1, m; x) \qquad p_V(x) = h_n(m, m-1; x).$$

Then $p_U(x)p_V(x) = [f_n(x)]^2$, so that $L(U+V) = L(X+Y)$ as desired.
Case 2: Suppose $n = 10$. Computation of

$$\begin{bmatrix} -S_{r-2}(x) & -S_{r-1} \\ S_{r-1}(x) & S_r(x) \end{bmatrix}^2 = \begin{bmatrix} 0 & -1 \\ 1 & x \end{bmatrix}^{2r} = \begin{bmatrix} -S_{2r-2}(x) & -S_{2r-1}(x) \\ S_{2r-1} & S_{2r}(x) \end{bmatrix}$$

yields the identity $S_{2r} = (S_r - S_{r-1})(S_r + S_{r-1})$. Taking $r = 2$ gives

$$S_4(2\cos\theta) = 0 \quad \text{iff} \begin{cases} (2\cos\theta)^2 + 2\cos\theta - 1 = 0 \\ \text{or} \\ (2\cos\theta)^2 - 2\cos\theta - 1 = 0 \end{cases}$$

Thus we may tabulate the values of $k_{10}(p) = 2\cos(2\pi p/10)$:

$$k_{10}(1) = (1 + \sqrt{5})/2 \qquad k_{10}(2) = (-1 + \sqrt{5})/2$$

$$k_{10}(3) = (1 - \sqrt{5})/2 \qquad k_{10}(4) = (-1 - \sqrt{5})/2.$$

(This explicit form for cos(36°) does not seem to be widely known.) A manageable

hand computation yields

$$
\begin{aligned}
h_{10}(3,4;x) &= (1+x)\left(1-k_{10}(1)x+x^2\right)\left(1-k_{10}(2)x+x^2\right)\left(1-k_{10}(3)x+x^2\right)^2\\
&= (x^9+1)+0\cdot(x^8+x)+(1/2)(1+\sqrt{5})(x^7+x^2)+0\cdot(x^6+x^3)+(x^5+x^4)\\
h_{10}(4,3;x) &= (1+x)\left(1-k_{10}(1)x+x^2\right)\left(1-k_{10}(2)x+x^2\right)\cdot\left(1-k_{10}(4)x+x^2\right)^2\\
&= (x^9+1)+2(x^8+x)+(1/2)\cdot(5-\sqrt{5})(x^7+x^2)+(3-\sqrt{5})(x^6+x^3)\\
&\quad +(4-\sqrt{5})(x^5+x^4).
\end{aligned}
$$

Each of these polynomials has non-negative coefficients. (Note that since there are some zero coefficients, machine approximations are not to be relied on for this example.)

Let U and V be 10-sided dice whose companion polynomials are

$$p_U(x) = h_{10}(3,4;x) \qquad p_V(x) = h_{10}(4,3;x).$$

As before, $L(U+V) = L(X+Y)$, but U and V are not fair.

Note. A computer search (using MACSYMA) revealed this as the only such example for $n \leqslant 11$.

REFERENCES

1. W. Feller, An Introduction to Probability Theory and Its Applications Vol. 2, J. Wiley, New York, 1966.
2. U. W. Hochstrasser, Orthogonal Polynomials, in Handbook of Mathematical Functions, (National Bureau of Standards Selected Government Publications) J. Wiley, New York, 1966.
3. I. Niven and H. Zuckerman, An Introduction to the Theory of Numbers, J. Wiley, New York, 1972.
4. T. Rivlin, The Chebyshev Polynomials, J. Wiley, New York, 1974.

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

Three Old Problems about Polynomials with Real Roots

RAPHAEL M. ROBINSON
Department of Mathematics, University of California, Berkeley, CA 94720

We consider only polynomials all of whose roots are real. For such a polynomial of degree $n > 1$, the **span** is the maximum distance between two roots, and the **gap** is the minimum distance. All three problems concern spans, and the third also involves gaps.

Problem 1. The polynomials considered here are monic and lie in $\mathbb{Z}[x]$, that is, have integer coefficients and leading coefficient 1. It was shown in [**1**] that there are infinitely many monic polynomials in $\mathbb{Z}[x]$ which are irreducible and have all their roots in $[a, b]$, provided that $b - a > 4$. Pólya had proved earlier that the number is finite if $b - a < 4$. If $b - a = 4$, then there are infinitely many such polynomials if a and b are integers. Are there any other cases with $b - a = 4$ where the number is infinite? Among the monic polynomials in $\mathbb{Z}[x]$ whose roots all lie in $[-2, 2.5]$ but not all in $[-2, 2]$, are there infinitely many which are irreducible and have span < 4? If the answer to the first question is yes, then so is the answer to the second, but perhaps not conversely.

Problem 2. Here we allow polynomials in $\mathbb{R}[x]$, that is, polynomials with real coefficients. An attempt was made in [**2**] to find the maximum possible span for the kth derivative of a polynomial $f(x)$ all of whose roots lie in $[-1, 1]$. The only nontrivial cases are those with $k + 2 \leqq n \leqq 2k + 1$. It was shown that in these cases the maximum can be attained only when all the roots of $f(x)$ lie at 1 or -1. The obvious conjecture is that these roots must be distributed as equally as possible between the two end points. How can this be proved?

Problem 3. A study was made in [3] of monic polynomials in $\mathbb{Z}[x]$ having real roots and span < 4. The objective was to determine all such polynomials of degree $n \leqq 8$ which are irreducible. In the process, it was necessary to compute some reducible polynomials with distinct roots. It was noticed that there is a strong correlation between reducibility and the smallness of the gap. For example, among the monic polynomials in $\mathbb{Z}[x]$ with span < 4, there are 17 essentially different sextics which are irreducible, and all have gaps > 0.23. On the other hand, there are 19 such sextics which are the product of two irreducible cubics, and 12 of these have gaps < 0.15. Is there a general theorem which determines values of n, s, and g, so that a monic polynomial in $\mathbb{Z}[x]$ of degree n with real roots is reducible if it has

span $< s$ and gap $< g$? Many reducible polynomials with distinct roots should satisfy the conditions. One suitable triple is $n = 6$, $s = 4$, and $g = 0.23$.

REFERENCES

1. Raphael M. Robinson, Intervals containing infinitely many sets of conjugate algebraic integers, Studies in Mathematical Analysis and Related Topics: Essays in Honor of George Pólya, Stanford University Press, Stanford, CA 1962, pp. 305–315.
2. ______, On the spans of derivatives of polynomials, this MONTHLY, 71 (1964) 504–508.
3. ______, Algebraic equations with span less than 4, *Math. Comp.*, 18 (1964) 547–559.

A Challenging Definite Integral

H. S. M. COXETER
Department of Mathematics, University of Toronto, Ontario, M5S 1A1, Canada

In spherical 3-space, that is, on the 3-sphere $x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1$ in Euclidean 4-space, a certain tetrahedron of known volume (which Schläfli called an **orthoscheme**) can be dissected into three smaller orthoschemes (two of them congruent) whose volumes can be expressed in terms of Schläfli functions [**2**, pp. 177–179; **1**, pp. vii, 6–12, 195]. It follows that

$$f + 2g = 2/15,$$

where

$$f = \frac{1}{\pi^2}\int_1^6 \frac{\operatorname{arc\,sec} t\,dt}{(t+2)\sqrt{(t+1)(t+3)}}, g = \frac{1}{\pi^2}\int_1^6 \frac{\operatorname{arc\,sec} t\,dt}{(t+2)\sqrt{t+1}}.$$

N. J. A. Sloane has used a computer to show that

$$f \approx 0.02268\ 05970\ 96406\ 8,$$

$$g \approx 0.05532\ 63681\ 18463\ 3.$$

The problem is to establish the precise value of $f + 2g$ without appealing to geometry or the computer!

REFERENCES

1. H. S. M. Coxeter, Twelve Geometric Essays, Southern Illinois University Press, Carbondale, IL, 1968.
2. Ludwig Schläfli, Gesammelte Mathematische Abhandlungen, II, Birkhäuser, Basel, 1953.

How Big a Slice Can You Make Through a Cube?

RAPHAEL M. ROBINSON
Department of Mathematics, University of California, Berkeley, CA 94720

Prove or disprove that for each $n \geq 2$, the maximum $(n-1)$-dimensional volume of a cross section of an n-dimensional unit cube is $\sqrt{2}$.

LETTERS TO THE EDITOR

Editor:

In the Notes section of this MONTHLY, Aug.–Sept. 1987, pp. 662–663, Boo Rim Choe offers an elegant proof of $\Sigma 1/n^2 = \pi^2/6$. It must, however, be remarked that this proof is by no means new: In a slightly different form it had been published by Euler in 1743 ([**1**]); cf. [**2**] (p. 388; footnote to nr. 210) and [**3**]. (Stäckel's paper, mentioned in [**3**], is reprinted in Euler's *Opera Omnia* I, 14, pp. 156–176.)

1. L. Euler, Demonstration de la somme de cette suite $1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \frac{1}{25} + \frac{1}{36} +$ etc., *Journal littéraire d'Allemagne*, 2:1 (1743) 115–127. (*Opera Omnia*, I, 14, 177–186.)
2. K. Knopp, Theorie und Anwendung der unendlichen Reihen, Grundlehren der mathematischen Wissenschaften 2, 5th ed., Springer-Verlag, 1964.
3. K. Knopp and I. Schur, Über die Herleitung der Gleichung $\Sigma_{n=1}^{\infty} 1/n^2 = \pi^2/6$, *Archiv der Math. und Physik*, 27 (1918) 174–176. (I. Schur, Gesammelte Abh., II, 246–248; Springer-Verlag 1973.)

Gerhard Turnwald
Universität Tübingen
Mathematisches Institut
Tübingen, West Germany

NOTES

EDITED BY DENNIS DETURCK, RICHARD LIBERA, AND ANITA E. SOLOW

Two Discrete Forms of the Jordan Curve Theorem

LAWRENCE NEFF STOUT
Department of Mathematics, Illinois Wesleyan University, Bloomington, IL 61702

The Jordan curve theorem is one of those frustrating results in topology: it is intuitively clear but quite hard to prove. In this note we will look at two discrete analogs of the Jordan curve theorem that are easy to prove by an induction argument coupled with some geometric intuition. One of the surprises is that when we discretize the plane we get two Jordan curve theorems rather than one, a consequence of the interplay between two natural products in the category of graphs. Topology in this context has been studied by Farmer in [**2**].

To state the discrete versions, we need to know what the discrete analog of the plane is and what plays the role of a simple closed curve. Since the plane is the topological product of two lines, we take as our discrete analog the product of two discrete lines. We will use undirected graphs for our analogs of spaces, with vertices for points and edges connecting points which are to be thought of as touching.

DEFINITION 1. A discrete n point line $[1, n]$ is a graph with vertices $\{1, 2, \dots, n\}$ and edges connecting each vertex to itself and to its successor. The discrete line L is a similar graph based on all of the integers.

DEFINITION 2. A discrete n point circle is a discrete n point line with n and 1 connected by an edge.

There are two important products in the category of graphs: the categorical product and the tight product. The tight product is used in building graphs using a sort of prime factorization in Behzad and Chartrand [**1**].

DEFINITION 3. The *product* of two graphs $(V_1, E_1)\prod(V_2, E_2)$ has the set $V_1 \times V_2$ as vertices and has (v_1, v_2) connected to (v_1', v_2') by the edge (e_1, e_2) if e_1 connects v_1 and v_1' and e_2 connects v_2 and v_2'.

DEFINITION 4. The *tight product* of two graphs has $V_1 \times V_2$ as its set of vertices and has an edge connecting (v_1, v_2) and (v_1', v_2') if and only if $v_1 = v_1'$ and there is an edge connecting v_2 and v_2', or $v_2 = v_2'$ and there is an edge connecting v_1 and v_1'. We denote this as $(V_1, E_1)\square(V_2, E_2)$.

If we take the product of two lines we get a patch of the plane with points connected which are nearest neighbors vertically, horizontally, or diagonally. If we take the tight product we leave out the diagonal connections.

The analog of continuous functions will be mappings of graphs: vertices are taken to vertices and edges to edges. A closed curve is the image of a circle under a graph map. It is simple if the map also reflects adjacency; that is, if $c(v)$ has an edge connecting it with $c(v')$ then v and v' had an edge connecting them too.

Simple curves then are forbidden to touch themselves, not just forbidden to cross themselves. This puts us in a position to state the two forms of the Jordan curve theorem.

THEOREM (Jordan curve theorem for tight closed curves). *If s is a simple closed curve with domain having at least* 8 *points in* $L \,\square\, L$, *then* $L \times L \setminus \mathrm{im}(s)$ *has exactly two product path components.*

THEOREM (Jordan curve theorem for product closed curves). *If s is a simple closed curve with domain having at least* 4 *points in* $L \sqcap L$, *then* $L \times L \setminus \mathrm{im}(s)$ *has exactly two tight path components.*

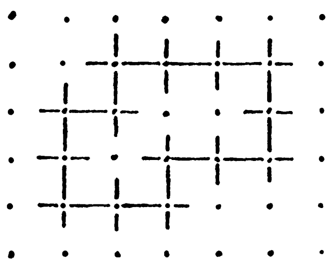


FIG. 1.

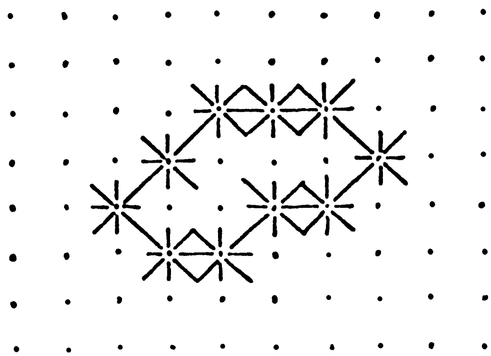


FIG. 2.

Notice that in Figure 1 the interior of the tight product closed curve is not connected in the tight product space. The interior is, however, connected in the product space, which allows diagonal connections. In the second illustration we have a simple closed curve in the product sense which fails to disconnect the categorical product space. If we use the tight product instead, then the interior is not connected to the exterior and each forms a connected set. The minimum size restriction eliminates the trivial cases in the next illustration.

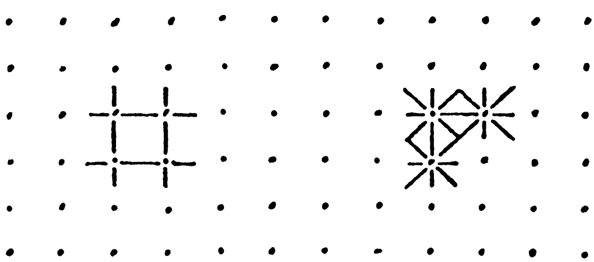


FIG. 3.

Proof (for product closed curves). Since a simple closed curve involves only a finite number of points we can move it into the first quadrant and guarantee that

the coordinates of points are bigger than 0 and less than m for sufficiently large m. We define the *rank* of s as the triple (N, X, Y) where N is the number of distinct points in the closed curve and (X, Y) is the point in the closed curve with largest first coordinate X and largest second coordinate Y of the points of $\mathrm{im}(s)$ with that first coordinate. Ranks are ordered lexicographically. This is a well-ordering, so strong induction on rank is a valid proof technique.

The smallest simple closed curve for this theorem has $N = 4$. It forms a diamond surrounding a single point which forms the inside component. All other points are connected to the point $(0,0)$ by a tight path. The requirement that a simple curve reflect adjacency eliminates other possible curves of length four. Thus the theorem is true for closed curves with length 4.

Now suppose that the theorem has been proved for all closed curves with rank less than (N, X, Y) and that s is a simple closed curve with rank (N, X, Y). We will reduce the rank by moving the point (X, Y) to $(X - 1, Y)$. The points in the closed curve s which were adjacent to (X, Y) could only be among $(X, Y - 1)$, $(X - 1, Y - 1)$, and $(X - 1, Y + 1)$. (Two points are adjacent to (X, Y) and they must be nonadjacent, hence, $(X - 1, Y)$ is not one of the possible points.) All of these are adjacent to $(X - 1, Y)$ so the result is still a closed curve, though it may not be a simple closed curve. Observe that moving this point reduces the rank. If the new closed curve is a simple closed curve then we are done since the interior of the original curve is the interior of the curve of lower rank with the point $(X - 1, Y)$, which is tight adjacent to it, added. The exterior of the original curve is the exterior of the new curve with the point (X, Y) removed. This is still tight connected since any tight path passing through (X, Y) in the exterior of the lower rank curve can take a detour through $(X, Y + 1)$, $(X + 1, Y + 1)$ and $(X + 1, Y)$.

There are two ways for the resulting closed curve to fail to be simple: either the point $(X - 1, Y)$ is adjacent to one of the points two steps away from (X, Y) in s, or it is adjacent to a point more than two steps away. If (X, Y) was $s(h)$ and $(X - 1, Y)$ is adjacent to $s(h - 2)$ then we can remove $s(h - 1)$. If (X, Y) was $s(h)$ and $(X - 1, Y)$ is adjacent to $s(h + 2)$ then we can remove $s(h + 1)$. Removing these points, if necessary, will further reduce the rank. The interior of the resulting curve is tight connected to (X, Y), so the interior of the original curve is tight connected. Any tight path passing through one of the points removed has a detour which avoids them and stays in the exterior. Figure 4 shows how this works for a typical case.

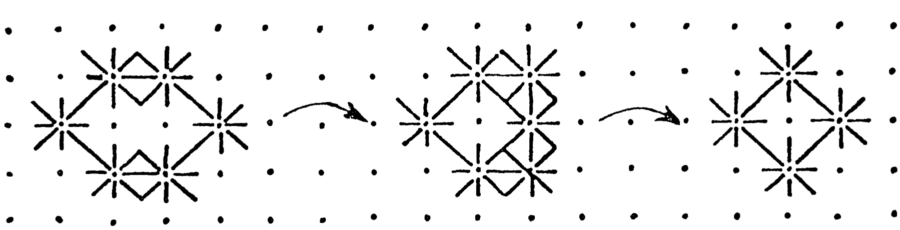


FIG. 4.

Suppose that (X, Y) is $s(h)$ and $(X - 1, Y)$ is adjacent to $s(k)$ where k is more than two away from h. Then by moving to $(X - 1, Y)$ we pinch the closed curve

into two closed curves which have a tight path connecting their interiors which passes through the point $(X-1, Y)$ and each of which is strictly shorter than our original loop. (See Figure 5.) Since they have smaller ranks they each divide the product into exactly two tight pieces. The interior of s is then the union of the interiors of these two new closed curves plus the point $(X-1, Y)$. It remains to show that the exterior is tight path connected.

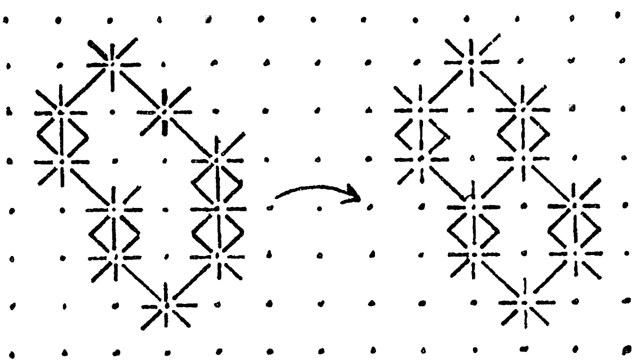


FIG. 5.

The exterior is the intersection of the exteriors of the two new closed curves. Call the new closed curves s_1 and s_2 and renumber so that the intersection points are at $t=0$ and $t=1$, with $(X-1, Y) = s_2(1)$. Let p and q be in $\text{ext}(s_1) \cap \text{ext}(s_2)$. Since $\text{ext}(s_1)$ is tight path connected there is a tight path in $\text{ext}(s_1)$ from p to q. If that path is also in $\text{ext}(s_2)$ then nothing more needs to be done. If not then there are points p' and q' such that p' is the last point in the path for which the segment from p to p' is in $\text{ext}(s_2)$ and q' is the first so that the segment from q' to q is in $\text{ext}(s_2)$. It follows that both p' and q' are adjacent to points in s_2. Thus to prove the theorem it will suffice to show that the set E of all points adjacent to s_2 and in the exterior of both curves is tight connected.

Since the original curve was simple we know that $s_2(0)$ and $s_2(1)$ are the only points in s_2 that are adjacent to points in s_1. We will show that E is tight path connected by walking around s_2 starting at (X, Y) and observing what happens in each nine point patch with an element of s_2 at the center. It is not difficult to list all of the ways that a product path can pass through a nine-point patch (see Figure 6) and in all cases the points on either side of the path form tight path connected sets. Since s_2 is of finite length we can piece together such patches to see that the set E is tight path connected.

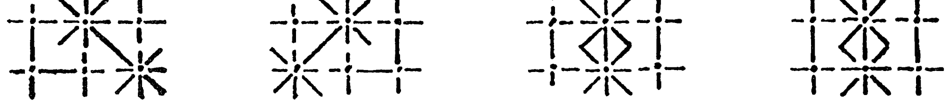


FIG. 6.

The proof of the theorem for simple closed curves in the tight product is a similar, though slightly less difficult, induction argument. The rank is defined the same way as in the product case. The reduction is done by moving the point (X, Y) to $(X - 1, Y - 1)$ which either gives a simple closed curve of lower rank or pinches the curve in two or gives a curve which can be shortened. Analysis of the possible nine-point patches again allows us to show that the exterior is connected.

The author would like to thank the referees and Randall Weiss for suggestions which improved this note.

REFERENCES

1. M. Behzad and G. Chartrand, Introduction to the Theory of Graphs, Allyn and Bacon, Boston, 1971.
2. F. Farmer, Homology of Reflexive Relations, *Math. Japonica*, V. 20 #1 (1975) 21–28.

Some Polynomial Identities that Imply Commutativity for Rings

MOHD. ASHRAF and MURTAZA A. QUADRI
Department of Mathematics, Aligarh Muslim University, Aligarh-202001, India

DANIEL ZELINSKY
Department of Mathematics, Northwestern University, Evanston, IL 60201

1. Introduction. Johnsen, Outcalt, and Yaqub [5] considered the ring-theoretic analogue of a well-known group-theoretic result, which states that a group G satisfying $(xy)^2 = x^2y^2$ for all x and y in G is necessarily abelian. It is also an easy exercise to show that a group G is abelian if $(xy)^2 = yx^2y$ for all x and y in G; the corresponding ring-theory version has not yet appeared in any text. It is surprising that the ring theoretic analogue of many such results escaped the attention of the research workers, though the commutativity of the rings satisfying other identities such as $xy^2x = yx^2y$ has been considered fully [1], [3], [6]. The reason for this sort of omission is understandable. We know that the essential mechanism in the proof of such results in groups is cancellation, which is not permissible in a general class of rings. Only a few results could be proved by going through several permutations of the substitutions of, say, y by $x + y$ and x by $x + 1$ starting with the given identity. To obtain other results, complicated combinatorial arguments had to be used [2], [4]. In this note our objective is to establish a result which allows a limited cancellation property in rings with unity. The proof depends on the simple strategy of substituting $x + 1$ for x to get another identity simpler than the original one. Indeed, we prove the following:

THEOREM A. *Let R be an associative ring with unity* 1 *and let $F(X, Y, Z)$ be a polynomial with coefficients from elements of R where the indeterminates commute neither with each other nor with the elements of R. Suppose that F is homogeneous in X of degree n and homogeneous in Y of degree m and that $F(x, y, xy - yx) = 0$ for all x and y in R. Then $m!n!F(1, 1, xy - yx) = 0$ for all x and y in R.*

In Section 3 we shall use our theorem to find some ring-theory version of the group-theoretic results including that of Johnsen, Outcalt, and Yaqub [**5**] and that of Awtar [**2**].

2. Proof of Main Theorem. Let $R[X, Y, Z]$ denote the ring of the polynomials in noncommuting indeterminates X, Y, Z over R. Define an automorphism σ on $R[X, Y, Z]$ by:

$$\sigma[F(X,Y,Z)] = F(X+1,Y,Z) \tag{1}$$

and a σ-derivation $\Delta = \sigma - I$:

$$\Delta[F(X,Y,Z)] = F(X+1,Y,Z) - F(X,Y,Z). \tag{2}$$

Easy computations show that for any two polynomials F and G in $R[X, Y, Z]$, we have

$$\Delta[F+G] = \Delta[F] + \Delta[G] \tag{3}$$

$$\Delta[FG] = (\Delta[F])(\sigma[G]) + F(\Delta[G]) \tag{4}$$

and an induction gives the Leibniz formula

$$\Delta^n[FG] = \sum_{r=0}^{n} \binom{n}{r}(\Delta^r[F])(\sigma^r\Delta^{n-r}[G]). \tag{5}$$

This allows us to prove:

LEMMA. *If F is homogeneous of degree n in X, then $\Delta^n[F(X, Y, Z)] = n!F(1, Y, Z)$ and $\Delta^m[F(X, Y, Z)] = 0$ for $m > n$.*

Proof of Lemma. By (3) it suffices to prove the lemma when $F(X, Y, Z)$ is a monomial. It can be proved by induction on n. If $n = 0$, $F(X, Y, Z)$ is independent of X and $\Delta^0[F(X, Y, Z)] = F(X, Y, Z) = F(1, Y, Z)$. Again $\Delta[F(X, Y, Z)] = F(X+1, Y, Z) - F(X, Y, Z) = 0$ and, hence, $\Delta^m[F(X, Y, Z)] = 0$ for all $m > n$.

For the induction step, write the monomial $F(X, Y, Z)$ as AXG, where A is a monomial with no X's in it, and G is a monomial of degree $n - 1$ in X. Then $\Delta[A] = 0$ and by (4), $\Delta[AX] = A$; hence, $\Delta^r[AX] = \Delta^{r-1}[A] = 0$ for $r > 1$ by the case $n = 0$. Again by using (5), we get

$$\Delta^m[AXG] = (AX)\Delta^m[G] + nA(\sigma\Delta^{m-1}[G]).$$

By the induction hypothesis, the first term on the right side is zero if $m \geqslant n$. The second term is zero if $m > n$; if $m = n$, this term equals $nA(n-1)!G(1, Y, Z) = n!F(1, Y, Z)$, which proves the lemma.

Proof of Theorem A. If $F(x, y, xy - yx) = 0$ for all x and y in R, then on replacing x by $x + 1$, we get

$$F(x+1, y, (x+1)y - y(x+1)) = F(x+1, y, xy - yx) = 0.$$

That is, if $F(x, y, xy - yx)$ is zero, the same is true for $\sigma[F]$, $\Delta[F]$, and $\Delta^n[F]$. Hence, $n!F(1, y, xy - yx) = 0$ for all x and y in R. By applying the whole

procedure again on the polynomial $F(1, Y, Z)$, which is homogeneous of degree m in Y by using a new derivation, Δ' defined as

$$\Delta'[F(X, Y, Z)] = F(1, Y + 1, Z) - F(Y, Z).$$

The result is the conclusion of the theorem.

3. Applications to Commutativity Theorems. We can derive a number of results with the help of our theorem proved above. As we have claimed in the beginning, even those results which could be proved earlier using very complicated combinatorial arguments will become corollaries of our theorem. We need just to select a suitable polynomial $F(X, Y, Z)$.

To begin with we prove the following result. Let us assume hence onward that R is an associative ring with unity 1.

PROPOSITION 1. *Let R be a ring satisfying $(xy)^2 = yx^2y$ for all x and y in R. Then R is commutative.*

Proof. Take $F(X, Y, Z) = ZXY$. Then indeed $F(x, y, xy - yx) = (xy)^2 - yx^2y = 0$. Hence, by applying Theorem A, $F(1, 1, xy - yx) = 0$, that is, $xy = yx$ and the ring R is commutative.

Similarly, by taking the polynomial $F(X, Y, Z) = XZY$ we prove the result due to Johnsen, Outcalt, and Yaqub **[5]**. As has been shown in example 3 of **[5]**, if we replace the identity $(xy)^2 = x^2y^2$ by $(xy)^3 = x^3y^3$, then the commutativity is not guaranteed. The following result suggests that in this case the commutativity fails only in rings that have no 2- or 3-torsion.

PROPOSITION 2. *Let R be a ring satisfying $(xy)^3 = x^3y^3$ for all x and y in R. If 6 is not a zero divisor in R, then R is commutative.*

Proof. Take $F(X, Y, Z) = X^2ZY^2 + XZXY^2 + XYXZY$, then $F(x, y, xy - yx) = x^3y^3 - (xy)^3 = 0$ and by our Theorem A, $n!m!F(1, 1, xy - yx) = 2!2!3(xy - yx) = 0$. But 2 and 3 are not zero divisors in R, so $xy - yx = 0$, which gives commutativity.

In fact, applying Theorem A we can derive even more general results. As an example we prove below a generalization of Proposition 2 which was earlier established by Awtar **[2]**.

PROPOSITION 3. *Let $n > 1$ be a positive integer and R be a ring in which no prime number $\leqslant n$ is a zero divisor. If R satisfies $(xy)^n = x^ny^n$ for all x and y in R, then R is commutative.*

Proof. Just as we did in the proof of Proposition 2, we take $F(X, Y, Z)$ to be a sum of $n(n - 1)/2$ monomials each of which is a product of one Z by $n - 1$ X's and $n - 1$ Y's (it takes one term to move each Y in $XYXY\ldots$ to the right, past one X). Then we have $F(x, y, xy - yx) = x^ny^n - (xy)^n = 0$, and by Theorem A,

$$m!n!F(1, 1, xy - yx) = ((n - 1)!)^2 \cdot (n(n - 1)/2)(xy - yx) = 0.$$

This implies that $xy - yx = 0$ if all primes dividing $((n - 1)!)^2(n(n - 1)/2)$ are not zero divisors.

REMARKS 1. If we take $F(X, Y, Z) = XZ - 2ZY$, the polynomial identity $(x + 2y)xy = xy(x + 2y)$, or if we take $F(X, Y, Z) = XYZ$, then the polynomial identity $(xy)^2 = xy^2x$ for all x and y in R implies commutativity. Other examples can be constructed ad libitum ad infinitum.

2. More subtle commutativity theorems, which do not work for all rings with unity, also often assume polynomial identities of the form $F(x, y, xy - yx) = 0$, but with $F(1, 1, Z) = 0$.

REFERENCES

1. Ram Awtar, A Remark on Commutativity of Certain Rings, *Proc. Amer. Math. Soc.*, 41 (1973) 370–372.
2. ______, On the Commutativity of Non-Associative Rings, *Publ. Math. (Debrecen)*, 22 (1975) 177–188.
3. R. N. Gupta, Nilpotent Matrices with Invertible Transpose, *Proc. Amer. Math. Soc.*, 24 (1970) 572–575.
4. A. Hermanci, Two Elementary Commutativity Theorems for Rings, *Acta Math. Acad. Sci. Hungar. Tomus*, 29 (1977) 23–29.
5. E. C. Johnsen, D. L. Outcalt, and Adil Yaqub, An Elementary Commutativity Theorem for Rings, *Amer. Math. Monthly*, 75 (1968) 288–289.
6. Murtaza A. Quadri, A Note on Commutativity of Semiprime Rings, *Math. Japonica*, 22 (1978) 509–511.

An Overlooked Example of Nonunique Factorization

HALE F. TROTTER
Department of Mathematics, Princeton University, NJ 08544

On the face of it, the familiar identity

$$\sin^2 t = 1 - \cos^2 t = (1 + \cos t)(1 - \cos t) \tag{1}$$

asserts that two different-looking pairs of factors have the same product. It seems to have gone unnoticed, however, that (1) is actually a valid example of nonunique factorization in an integral domain when looked at in the proper context. Its familiarity makes it a particularly attractive example to present to students encountering nonunique factorization for the first time. Just as the usual textbook examples involving integers in quadratic number fields, such as $2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5})$, show that unique factorization can fail in rings very much like the integers, the example treated here shows that it can fail in a ring very much like the ring of polynomials over a field.

Of course we have to do more than simply remark that the two sides of (1) look different. We must specify the ring we are working in, and then show that the factors $\sin t$, $1 + \cos t$, and $1 - \cos t$ are irreducible, and that $\sin t$ is not the product of one of the other factors and a unit (invertible element) of the ring.

We shall work with the real trigonometric polynomials, that is, the functions representable as finite sums of the form

$$a_0 + \sum^{k} (a_n \cos nt + b_n \sin nt) \tag{2}$$

in which the a's and b's are real numbers. Students who have seen anything of Fourier series find it natural enough to consider these functions, although they may not have seen them called trigonometric polynomials, or considered the question of whether they form a ring. The familiar Fourier coefficient formulas $a_0 = (2\pi)^{-1}\int_{-\pi}^{\pi} f(x)\,dx$, $a_n = \pi^{-1}\int_{-\pi}^{\pi} f(x)\cos nx\,dx$, and $b_n = \pi^{-1}\int_{-\pi}^{\pi} f(x)\sin nx\,dx$ for $n > 0$, show that the coefficients in (2) are uniquely determined by the function.

The *degree* of a nonzero trigonometric polynomial is defined as the largest value of n for which a_n and b_n are not both zero. The following well-known lemma shows that the trigonometric polynomials form a ring, and that degrees behave as they do for ordinary polynomials.

LEMMA. *The product of a trigonometric polynomial of degree m and one of degree n is a trigonometric polynomial of degree $m + n$.*

Proof. The assertion of the lemma is obvious if m or n is 0, because a trigonometric polynomial of degree zero is simply a constant function. From now on we assume $m, n > 0$. Recall the standard identities for expressing products of sines and cosines in terms of sums and differences of other sines and cosines:

$$(\sin a)(\sin b) = [\cos(a - b) - \cos(a + b)]/2$$

$$(\cos a)(\cos b) = [\cos(a - b) + \cos(a + b)]/2$$

$$(\sin a)(\cos b) = [\sin(a + b) + \sin(a - b)]/2.$$

Applying these to the product of $p\cos mt + q\sin mt$ and $r\cos nt + s\sin nt$ and collecting terms, gives the result

$$A\cos(m - n)t + B\sin(m - n)t + C\cos(m + n)t + D\sin(m + n)t, \tag{3}$$

where $A = (pr + qs)/2$, $B = (ps - qr)/2$, $C = (pr - qs)/2$, and $D = (ps + qr)/2$. When $m > n$, (3) is already in the form (2). If $n > m$, replacing $\cos(m - n)$ by $\cos(n - m)$ and $\sin(m - n)$ by $-\sin(n - m)$ puts it in the proper form, while if $m = n$ it is necessary to replace $\sin 0$ by 0 and $\cos 0$ by 1. Direct calculation gives

$$C^2 + D^2 = (p^2 + q^2)(r^2 + s^2)/4,$$

which shows that if neither factor is zero (so $p^2 + q^2 \neq 0$ and $r^2 + s^2 \neq 0$) then $C^2 + D^2 \neq 0$, so the product has degree $m + n$.

Now consider the product of any two trigonometric polynomials of respective degrees m and n. It is a sum of products of terms of the type just considered, so it is a trigonometric polynomial. The product of the two high-order terms gives a non-zero term of degree $m + n$ which cannot be cancelled by any other term in the product, so the result has degree $m + n$ as claimed.

PROPOSITION. *The trigonometric polynomials form an integral domain. Furthermore,*

(a) *The units (invertible elements) in this domain are the elements of degree* 0, *that is, the constant functions.*
(b) *All elements of degree* 1, *including* $\sin t$, $1 + \cos t$, *and* $1 - \cos t$, *are irreducible.*

The proposition follows at once from the lemma, just as with ordinary polynomials, and we leave the details to the reader.

It follows from (a) and (b) that the factors in (1) are irreducible and that $\sin t$ is not the product of one of the other factors with a unit. Hence we have a genuine case of nonunique factorization.

One can very well stop here in an elementary discussion, but the example does raise another point that may be of interest.

The proof of the lemma uses the fact that the sum of the squares of two real numbers is zero only when both are zero, and breaks down if complex coefficients are allowed. Using the complex exponential forms of the sine and cosine shows that the ring of trigonometric polynomials with complex coefficients is the same as the ring of polynomials in positive and negative powers of $z = e^{it}$ with complex coefficients. To see that this is a unique factorization ring, define the degree of a polynomial in z and z^{-1} as the difference between the largest and smallest exponents appearing in non-zero terms. With this definition, the elements of degree zero are the monomials, which are exactly the invertible elements in this ring. The usual proof that ordinary polynomials over a field form a Euclidean ring then goes through with no essential change.

What is it about the change of coefficients that alters the nature of factorization in the ring? For one thing, introducing complex coefficients produces many more units—all the non-zero constant multiples of powers of $z = \cos t + i \sin t$ and $z^{-1} = \cos t - i \sin t$. Our particular example breaks down because the factors involved cease to be irreducible. We have

$$\sin t = (z - z^{-1})/(2i) = z^{-1}(z - 1)(z + 1)/(2i),$$

$$1 - \cos t = (-z + 2 - z^{-1})/2 = -z^{-1}(z - 1)^2/2,$$

and

$$1 + \cos t = z^{-1}(z + 1)^2/2,$$

so both sides of (1) become

$$-z^{-2}(z - 1)^2(z + 1)^2/4$$

when expressed as a product of irreducible factors.

A ring of algebraic integers can sometimes be enlarged to another in a way that restores unique factorization, although the problem of how and when it can be done is not at all elementary, and as far as I know is not solved in general. For example, the ring $\mathbf{Z}[\sqrt{-3}]$ consisting of numbers of the form $a + b\sqrt{-3}$ with a and b integers does not have unique factorization, as the equation $2 \cdot 2 = (1 + \sqrt{-3})(1 - \sqrt{-3})$ shows. Unique factorization can be restored in this case by enlarging to the ring of *all* algebraic integers in the field $\mathbf{Q}(\sqrt{-3})$, which is $\mathbf{Z}[\omega]$, where $\omega = (1 + \sqrt{-3})/2$ is a complex cube root of one. This does not work for the ring $\mathbf{Z}[\sqrt{-5}]$ used in the example at the beginning of the paper, because that is already the ring of all algebraic integers in the field $\mathbf{Q}(\sqrt{-5})$. The ring of all algebraic integers in the enlarged field $\mathbf{Q}(\sqrt{-5}, i)$, however, which can be shown to be the ring $\mathbf{Z}[\eta]$ where $\eta = (i + \sqrt{-5})/2$ is a root of $x^4 + 3x^2 + 1$, is an enlargement of $\mathbf{Z}[\sqrt{-5}]$ that does have unique factorization. I do not know an elementary proof of

the last assertion, but it is easily established by standard arguments based on Minkowski's estimate, as illustrated in [**1**, chapter 12], [**2**, chapter 13], or [**3**, chapter 5].

REFERENCES

1. E. Artin, Theory of Algebraic Numbers, Math. Inst. Göttingen, 1959.
2. H. Cohn, A Classical Invitation to Algebraic Numbers and Class Fields, Springer, New York, 1978.
3. S. Lang, Algebraic Number Theory, Addison-Wesley, Reading, 1970.

Two Counterexamples in General Topology

P. V. KRISHNAIAH
Department of Mathematics, The University of Toledo, Ohio

In [**3**] Albert Wilansky inserted two axioms between the separation axioms T_1 and T_2, namely the US-axiom (every convergent sequence has a unique limit) and the KC-axiom (every compact subset is closed). He showed that $T_2 \Rightarrow KC \Rightarrow US \Rightarrow T_1$ and discussed at length the problem of constructing counterexamples of compact spaces showing the failure of each of the reverse implications, proving several interesting results in the process. As a consequence of theorems 4 and 5 of [**3**], it was brought out that, if a T_2-space X is not a k-space (a k-space is one in which a set is closed iff its intersection with each closed compact set is closed) then its one point compactification X^+ is US but not KC. An example of a T_2-space, which is not a k-space is given in [**3**], example 7. Since this example involves the Cech compactification, it seems worthwhile to have the following two elementary examples.

Example 1. The Appert space A (see [**1**], p. 117) whose ground set is the set of all positive integers and $E \subseteq A$ is open iff either $1 \notin E$ or $1 \in E$ and

$$\lim_{n\to\infty} \frac{1}{n}\left\{\sum_{r=1}^{n} \chi_E(r)\right\} = 1.$$

Example 2. The Fortissimo space F (see [**1**], p. 53) whose ground set is any uncountable set with a particular point p and $E \subseteq F$ is open iff either $p \notin E$ or $p \in E$ and $F \setminus E$ is countable.

One easily verifies that both A and F above are noncompact T_2-spaces, that both are pseudofinite (i.e., all compact subsets are finite) and that neither is discrete. Hence, neither of them is a k-space so that both A^+ and F^+ are US but not KC. This can in fact be shown directly. While $A^+ \setminus \{1\}$ and $F^+ \setminus \{p\}$ are compact nonclosed subsets of A^+ and F^+, respectively, one easily imitates the proof of theorem 4 in [**3**] to show that they are US spaces.

Other examples may be found in [**2**, example 2.3] and [**4**, p. 345].

REFERENCES

1. Lynn A. Steen and J. Arthur Seebach Jr., Counterexamples in Topology, Holt, Rinehart and Winston, New York, 1970.
2. N. E. Steenrod, A Convenient Category of Topological Spaces, *Michigan Math. J.*, 14 (1967) 133–152.
3. Albert Wilansky, Between T_1 and T_2, *Amer. Math. Monthly*, 74 (1967), No. 3, pp. 261–266.
4. ______, Topology for Analysis, Ginn and Company, Waltham, Massachusetts, 1970.

THE TEACHING OF MATHEMATICS

EDITED BY JOAN P. HUTCHINSON AND STAN WAGON

An Elementary Approach to $y'' = -y$

J. L. BRENNER,
10 Phillips Rd., Palo Alto, CA 94303

THEOREM. *Every solution of $y'' + k^2y = 0$ $(k > 0)$ on the interval $a < x < b$ has the form $y = c_1\cos kx + c_2\sin kx$, where c_1, c_2 are constants.*

Proof. Without loss of generality, take $k = 1$ (the general case comes from the change of independent variable $t = kx$). Define

$$\begin{aligned} c_1 &= y\cos t - y'\sin t, \\ c_2 &= y\sin t + y'\cos t. \end{aligned} \tag{1}$$

Since $y'' + y = 0$, it is clear that $c_1' = c_2' = 0$, so that c_1, c_2 are constant. Eliminate y' from equations (1). □

There are many other proofs of this theorem. From $y'(y'' + y) = 0$, one finds $y'^2 + y^2 = r^2 = \text{const}$, so that (if $r > 0$) $dy/\sqrt{r^2 - y^2} = \pm dt$, $y = r\sin(\pm t + C)$. See also [**1**, pp. 83–84].

REFERENCE

1. P. G. Kumpel and J. A. Thorpe, Linear Algebra, Saunders, Philadelphia, 1983.

A Simple Proof for the Simplicity of A_5

BENNO ARTMANN
Fachbereich Mathematik, TH Darmstadt, Schloßgartenstr. 7, D-6100 Darmstadt, West Germany

Dedicated to Günther Pickert on the occasion of his 70th birthday

The alternating group A_5 is usually the first example of a simple nonabelian group for the student. The group A_5 is isomorphic to the group D of rotations of a regular dodecahedron (or icosahedron). I think it may be worthwhile to see a proof for the simplicity of D whose ideas are *immediately transparent*, *easily remembered*, and use nothing more than the concept of a group and a homomorphism between

PROBLEMS AND SOLUTIONS

DEDICATED TO THE MEMORY OF ISRAEL N. HERSTEIN

For instructions about submitting ***proposed*** *problems for publication in this department see the inside front cover. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

Solutions *should be sent to the address given on the inside front cover. Two copies suffice.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

For instructions about submitting solutions of Problems, which should be mailed before August 31, 1988, see the inside front cover. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label if an acknowledgment is desired.

E 3259. *Proposed by Jordi Dou, Barcelona, Spain.*

Let R be a semicircular region bounded by a line L and a semicircle S with center on L. Suppose P_1 and P_2 are given points in the interior of R. We wish to find parallel lines l_1, l_2 through P_1, P_2, respectively, such that

$$\frac{P_1C_1}{P_1D_1} = \frac{P_2C_2}{P_2D_2},$$

where C_1, D_1 are the intersections of l_1 with S and L and C_2, D_2 are the intersections of l_2 with S and L. Give a necessary and sufficient condition on P_1, P_2 for such parallel lines to exist.

E 3260. *Proposed by Peter Andrews and Edward T. H. Wang, Wilfrid Laurier University, Waterloo, Ontario, Canada.*

In how many ways can n white squares and n black squares be chosen from a $2n$ by $2n$ chessboard in such a way that no two of the chosen squares lie in the same row or the same column?

E 3261. *Proposed by Detlef Laugwitz, Technische Hochschule, Darmstadt, West Germany.*

Let G be the group of 3 by 3 orthogonal matrices with rational entries. Let S be the subset of G consisting of the six 3 by 3 permutation matrices and all matrices of the form

$$\begin{pmatrix} \frac{a}{c} & \frac{b}{c} & 0 \\ -\frac{b}{c} & \frac{a}{c} & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

where a, b, c are integers with $a^2 + b^2 = c^2 > 0$. Does S generate G?

E 3262. *Proposed by R. G. E. Pinch, Emmanuel College, Cambridge and C. Boyd, University of Edinburgh.*

It is known that every natural number n can be expressed as the sum of four squares of integers, and that three squares suffice unless n is of the form $4^a(8b+7)$. Show that every sufficiently large natural number can be expressed as the sum of at most five squares of composite numbers (i.e., squares of positive integers divisible either by the square of a prime or by the product of two distinct primes).

E 3263. *Proposed by Walther Janous, Ursulinengymnasium, Innsbruck, Austria.*

For $n \geqslant 2$ let

$$H_n = \{X = (x_1, \ldots, x_n) \colon x_1 + \cdots + x_n = 1, \quad x_1 > 0, \ldots, x_n > 0\}.$$

For $1 \leqslant k \leqslant n$ and $X \in H_n$ let

$$S_k(X) = \sum \frac{x_{i_1}}{1 - x_{i_1}} \cdots \frac{x_{i_k}}{1 - x_{i_k}},$$

where the sum extends over all k-tuples $(i_1, \ldots, i_k)$ with $1 \leqslant i_1 < \cdots < i_k \leqslant n$. Put

$$M_k(n) = \sup_{X \in H_n} S_k(X).$$

It is not hard to see that $M_n(n) = (n-1)^{-n}$ (D. S. Mitrinović, *Analytic Inequalities*, Springer, Berlin, 1970, (3.2.46) on p. 214).

(a) Show that $M_2(n) = 1$.

(b) Show that $M_3(4) = 4/27$.

*(c) For what pairs k, n with $3 \leqslant k \leqslant n$ is it true that $M_k(n) = \binom{n}{k}(n-1)^{-k}$?

E 3264. *Proposed by Stanley Rabinowitz, Alliant Computer Systems Corp, Littleton, MA.*

Let P_n/Q_n be the nth convergent for the continued fraction

$$\frac{1}{1+}\,\frac{1}{2+}\,\frac{1}{3+}\,\frac{1}{4+}\cdots,$$

i.e., let $P_1 = 1$, $Q_1 = 1$, $P_2 = 2$, $Q_2 = 3$, and

$$P_n = nP_{n-1} + P_{n-2}, \qquad Q_n = nQ_{n-1} + Q_{n-2} \qquad (n \geqslant 3).$$

(Cf. Hardy and Wright, An Introduction to the Theory of Numbers, Chapter 10.)

Give asymptotic estimates for P_n and Q_n.

SOLUTIONS OF ELEMENTARY PROBLEMS

Maximal Minimal Card Shufflings

E 3143 [1986, 299]. *Proposed by Allen J. Schwenk, Western Michigan University, Kalamazoo, MI.*

A riffle shuffle of a deck of cards is the commonly used technique of cutting the deck into two portions (not necessarily equal), then, elevating the corners slightly, allowing each portion to fall card by card (not necessarily alternating) merging with the other portion, and finally pushing them together to reconstitute the pack. Given a deck of n cards in arbitrarily permuted order π, determine as a function of π the minimum number of riffle shuffles that could possibly produce the identity sequence $1, 2, \ldots, n$. Describe a procedure that attains this minimum. Which original sequences require the most shuffles?

Solution by the proposer. For the permutation $\pi = a_1, \ldots, a_n$, let a *descent* be a position j such that $a_j > a_{j+1}$. By convention, we always consider position n to be a descent; thus the identity permutation is the only sequence with 1 descent. We show that the minimum number of shuffles required is $\lceil \log_2 d(\pi) \rceil$, where $d(\pi)$ is the total number of descents in π.

By partitioning the sequence π at the descent positions, we may view π as consisting of $d = d(\pi)$ blocks $B_1, \ldots, B_d$ that are strictly increasing. Now perform a riffle shuffle by cutting the deck after the block $B_{\lceil d/2 \rceil}$ and merging the two portions so that block B_j merges with block $B_{\lceil d/2 \rceil + j}$ to form a new increasing block C_j, for each $j \leqslant \lfloor d/2 \rfloor$. If d is odd, the last block $B_{\lceil d/2 \rceil}$ is unaltered. Thus, one shuffle transforms π to a sequence with $\lceil d/2 \rceil$ descents, and iterating this procedure produces the identity permutation after $\lceil \log_2 d(\pi) \rceil$ shuffles.

On the other hand, any cut of the original deck leaves at least $\lceil d/2 \rceil$ descents in one portion of the deck, and merging cannot reduce this number. By induction on d, at least $\lceil \log_2 d(\pi) \rceil$ shuffles are needed to complete the job, so $\lceil \log_2 d(\pi) \rceil$ are needed for a sequence with d descents.

The maximum required number of shuffles is $k = \lceil \log_2 n \rceil$, and the permutations requiring this many shuffles are those with more than 2^{k-1} descents.

Also solved by K. Schilling and by L. Szuecs.

Generating Sets in a Topological Space

E 3144 [1986, 299]. *Proposed by Edwin Buchman, California State University, Fullerton, CA.*

Determine the maximum number of sets in a topological space which can be generated from one set by application (possibly repeated, in any order) of the operations of taking the complement, interior, and boundary of a set.

Solution by Jesús Ferrer, Cami Collado, Oliva (Valencia), Spain. At most 34 different sets can be constructed from one set by repeatedly taking the complement, interior, or boundary. Let A', A^0, A^* denote the complement, interior, and boundary of a set A. Then the 34 sets that can be constructed appear in the following tree.

A	A^*	A^{**}	$A^{**\prime}$			
		A^{*0}	A^{*0*}	$A^{*0*\prime}$		
			$A^{*0\prime}$	$A^{*0\prime 0}$	$A^{*0\prime 0\prime}$	
		$A^{*\prime}$				
	A^0	A^{0*}	$A^{0*\prime}$			
		$A^{0\prime}$	$A^{0\prime 0}$	$A^{0\prime 0*}$	$A^{0\prime 0*\prime}$	
				$A^{0\prime 0\prime}$	$A^{0\prime 0\prime 0}$	$A^{0\prime 0\prime 0\prime}$
	A'	$A^{\prime 0}$	$A^{\prime 0*}$	$A^{\prime 0*0}$	$A^{\prime 0*0\prime}$	
				$A^{\prime 0*\prime}$		
			$A^{\prime 0\prime}$	$A^{\prime 0\prime 0}$	$A^{\prime 0\prime 0*}$	$A^{\prime 0\prime 0*\prime}$
					$A^{\prime 0\prime 0\prime}$	$A^{\prime 0\prime 0\prime 0}$	$A^{\prime 0\prime 0\prime 0\prime}$

The lack of additional sets follows from $A^* = A^{\prime 0\prime} \cap A^{0\prime}$, $A'^* = A^*$, $A^{*\prime 0} = A^{*\prime}$, $A^{**0} = \emptyset = A^{0*0}$, $A^{0**} = A^{0*}$, $A^{\prime 0\prime 0\prime 0} = A^{\prime 0\prime 0}$, $A^{0\prime 0\prime 0\prime 0} = A^{0\prime 0}$, $A^{\prime 0\prime 0\prime 0*} = A^{\prime 0\prime 0*}$, $A^{0\prime 0\prime 0*} = A^{0\prime 0*}$, $A^{*0\prime 0\prime 0} = A^{*0}$, $A^{*0\prime 0*} = A^{*0*}$, $A^{***} = A^{**}$. Also $A^{\prime 0*0} = \emptyset$ and $A^{\prime 0*0\prime} = X$.

To see that the number 34 cannot be reduced, let X be the real numbers with the usual topology, and let $\mathbb{Q}$ be the rationals. Then all 34 sets are distinct when A is the set

$$A = [0,1) \cup (1,2) \cup \{3\} \cup (\mathbb{Q} \cap [4, \infty)).$$

Editorial comment. This problem is very similar to Advanced Problem 5996, [1974, 1034; 1978, 283]. That problem asks the same question for {closure, interior, union}, the answer being that at most 13 distinct sets can be constructed (see also Closure, interior, and union in finite topological spaces, *Colloq. Math.*, 38 (1977) 41–51, by L. E. Moser). Kuratowski showed that at most 14 distinct sets can be constructed using {closure, interior, complement} (see Sur l'opération $\bar{A}$ de l'analyses situs, *Fund. Math.*, 3 (1922) 182–199.

Also solved by M. Bowron, J. Hook, O. P. Lossers (Netherlands), O. Matouš (Czechoslovakia), W. D. McIntosh, L. F. Meyers, R. Patenaude, J. P. Robertson, and the proposer.

An Integral of Cosines

E 3145 [1986, 299]. *Proposed by Clinton J. Kolaski, University of Minnesota, Duluth.*

Show that

$$\int_0^{\pi} \frac{\cos nx - \cos ny}{\cos x - \cos y}\, dx = \pi \frac{\sin ny}{\sin y} \qquad (n = 0, 1, 2, \dots).$$

Solution I by W. O. Egerland and C. E. Hansen, University of Baltimore. Put

$$f_n(x, y) = \frac{\cos nx - \cos ny}{\cos x - \cos y}$$

for $n = 0, 1, 2, \dots$ and $(x, y) \in R^2$. Since there is a polynomial T_n of degree n such that $\cos n\theta = T_n(\cos\theta)$, it follows that $f_n(x, y)$ may be expressed as a polynomial in $\cos x$ and $\cos y$ of degree $n - 1$. Clearly $f_0(x, y) = 0$, $f_1(x, y) = 1$, $f_2(x, y) = 2\cos y + 2\cos x$. The addition formula for the cosine immediately gives

$$\begin{aligned} f_{n+1}(x, y) + f_{n-1}(x, y) &= \frac{2\cos nx \cos x - 2\cos ny \cos y}{\cos x - \cos y} \\ &= 2\cos nx + 2f_n(x, y)\cos y \end{aligned} \tag{1}$$

for $n = 1, 2, \dots$. In view of the identity

$$\sin(r + 1)y + \sin(r - 1)y = 2\sin ry \cos y,$$

a straightforward induction argument shows that the recurrence (1) and the above initial values of $f_n(x, y)$ imply the following explicit formula

$$f_n(x, y) = \frac{\sin ny}{\sin y} + 2\sum_{k=1}^{n-1} \cos kx \frac{\sin(n - k)y}{\sin y} \qquad (n = 2, 3, \dots). \tag{2}$$

Thus for fixed y we have the indefinite integral

$$\int f_n(x, y)\, dx = x\frac{\sin ny}{\sin y} + 2\sum_{k=1}^{n-1} \frac{\sin kx}{k} \frac{\sin(n - k)y}{\sin y} \qquad (n = 2, 3, \dots). \tag{3}$$

The result of the problem follows.

Solution II by Kwang Kyu Park, Korea Advanced Institute of Science and Technology, Seoul, Korea. The following formulas are well known

$$\sum_{n=0}^{\infty} r^n \cos nx = \frac{1 - r\cos x}{1 - 2r\cos x + r^2} \qquad (|r| < 1), \tag{4}$$

$$\int_0^{\pi} \frac{dx}{1 - 2r\cos x + r^2} = \frac{\pi}{1 - r^2} \qquad (|r| < 1), \tag{5}$$

$$\sum_{n=0}^{\infty} r^n \sin ny = \frac{r\sin y}{1 - 2\cos y + r^2} \qquad (|r| < 1). \tag{6}$$

Since

$$\frac{\cos nx - \cos ny}{\cos x - \cos y} \leqslant n^2$$

for $(x, y) \in R^2$, the series

$$\sum_{n=0}^{\infty} \frac{\cos nx - \cos ny}{\cos x - \cos y} r^n$$

converges uniformly in x, y provided $|r| < 1$. Thus, if $|r| < 1$, we have by (4)

$$\sum_{n=0}^{\infty} r^n \int_0^{\pi} \frac{\cos nx - \cos ny}{\cos x - \cos y}\, dx$$

$$= \int_0^{\pi} \sum_{n=0}^{\infty} r^n \frac{\cos nx - \cos ny}{\cos x - \cos y}\, dx$$

$$= \int_0^{\pi} \left\{ \frac{1 - r\cos x}{1 - 2r\cos x + r^2} - \frac{1 - r\cos y}{1 - 2r\cos y + r^2} \right\} \frac{dx}{\cos x - \cos y}$$

$$= \int_0^{\pi} \frac{r - r^3}{1 - 2r\cos y + r^2} \frac{dx}{1 - 2r\cos x + r^2}.$$

Using (5) and (6) in turn, we obtain

$$\sum_{n=0}^{\infty} r^n \int_0^{\pi} \frac{\cos nx - \cos ny}{\cos x - \cos y}\, dx = \frac{r - r^3}{1 - 2r\cos y + r^2} \frac{\pi}{1 - r^2} = \sum_{n=0}^{\infty} \pi \frac{\sin ny}{\sin y} r^n.$$

Comparing coefficients of r^n, we get the desired formula.

Solution III by Kee-wai Lau, Hong Kong. Denote the integral of the problem by I_n. Substituting $z = e^{ix}$ we have

$$2I_n = \int_{-\pi}^{\pi} \frac{\cos nx - \cos ny}{\cos x - \cos y}\, dx = -i \int_c \frac{z^n + z^{-n} - (e^{iny} + e^{-iny})}{z + z^{-1} - (e^{iy} + e^{-iy})} \frac{dz}{z},$$

where c is the positive orientation of the unit circle. It follows that

$$2I_n = -i \int_c \frac{(z^n - e^{iny})(z^n - e^{-iny})}{(z - e^{iy})(z - e^{-iy})} \frac{dz}{z^n}$$

$$= -i \int_c \sum_{k=1}^{n} z^{n-k} e^{i(k-1)y} \sum_{r=1}^{n} z^{n-r} e^{-i(r-1)y} z^{-n}\, dz.$$

The coefficient of z^{-1} in the last integrand is

$$\sum_{k=1}^{n} e^{i(k-1)y} e^{-i(n-k)y} = e^{-(n+1)iy} \sum_{k=1}^{n} e^{2kiy}.$$

By the residue theorem $I_n = \pi \sin ny/\sin y$ if y is not an integral multiple of π and

$I_n = n\pi(-1)^{m(n+1)}$ if $y = m\pi$. This calculation could also be carried out in terms of the original variable of integration x, using the orthogonality of the functions $e^{inx}(n = 0, \pm 1, \pm 2, \ldots)$ over $[-\pi, \pi]$ instead of the residue theorem.

Editorial Comment. S. K. Ntouyan and Hang-Fai Yeung (Australia) each pointed out that the indefinite integral (3) occurs as a problem on page 205 of Joseph Edwards, *A Treatise of the Integral Calculus*, Volume 1, Macmillan, London, 1921. Edwards attributes (3) to Hermite.

Integrating (1) (with respect to x) over $[-\pi, \pi]$ gives

$$I_{n+1} - 2I_n \cos y + I_{n-1} = 0.$$

Most solvers based their solutions on this recurrence formula and the initial values $I_0 = 0$ and $I_1 = \pi$, using either mathematical induction or the theory of second-order linear difference equations with constant coefficients. M. S. Klamkin, Morris Morduchow, and R. A. Struble observed that this method of solving the problem occurs explicitly on pages 92–93 of H. Glauert, *The Elements of Aerofoil and Airscrew Theory*, Cambridge University Press, Cambridge, 1947, and on page 80 of L. M. Milne-Thomson, *Theoretical Aerodynamics*, Macmillan, London, 1952.

If λ is any positive real number and $0 < y < \pi$, the quotient

$$\frac{\cos \lambda x - \cos \lambda y}{\cos x - \cos y} = \frac{\sin\{\lambda(x+y)/2\}\sin\{\lambda(x-y)/2\}}{\sin\{(x+y)/2\}\sin\{(x-y)/2\}}$$

is bounded for $0 \leqslant x \leqslant \pi$ and thus the integral

$$I_\lambda = \int_0^\pi \frac{\cos \lambda x - \cos \lambda y}{\cos x - \cos y}\, dx$$

exists. In *Siam Review* 24 (1982), 83–85, Solution of Problem 81-5, J. A. Boa shows that

$$I_\lambda = \frac{\pi \sin \lambda y}{\sin y} - \frac{2 \sin \lambda \pi}{\sin y} \sum_{n=0}^{\infty} \frac{(-1)^n \sin(n+1) y}{n + \lambda + 1}.$$

If λ is rational, the series here can be summed in finite form. Cf. also *Siam Review*, 29 (1987) 303–305, Solution of Problem 86-10.

Solved also by 35 other readers and the proposer.

The Countable Co-countable Algebra

E 3147 [1986, 400]. *Proposed by A. Wilansky, Lehigh University, Bethlehem, PA.*

Let (X, B, μ) be a measure space such that all singletons are measurable. Let $f(x) = \mu(\{x\})$. Must f be measurable?

Solution by G. Turnwald, Mathematisches Institut der Universität, Tübingen, West Germany. The answer is No. Let X be the disjoint union of uncountable sets X_1 and X_2, and let B be the σ-algebra of subsets A of X such that A or $X - A$ is countable. Let $\mu(A) = |A \cap X_1|$, if $A \cap X_1$ is finite, and otherwise $\mu(A) = \infty$. Since $f(x) = 1$ on X_1 and $f(x) = 0$ on X_2, $f^{-1}(1)$ and $f^{-1}(0)$ are uncountable. Hence f is not measurable.

Editorial comment. Several solvers noted that the answer is Yes if $\mu(X)$ is finite, since then $f(x) = 0$ except on a countable set, which implies the measurability of f.

Also solved by J. Ferrer (Spain), J. A. Goldstein, E. Hertz, A. A. Jagers (Netherlands), D. H. King, R. Levy, J. Ling and P. G. Walsh (Canada), O. P. Lossers (Netherlands), L. A. Lucas, D. Neuenschwander (student, Switzerland), Oxford Running Club (University of Mississippi), V. Pambuccian (Romania), E. Posti (Finland), D. Ramachandran, R. H. Scissors, A. V. Stanoyevitch, J. C. Tripp, and the proposer.

Matching Socks

E 3148 [1986, 400]. *Proposed by Rick Luttmann, Sonoma State University, Rohnert Park, CA.*

Let n distinct pairs of socks be put into the laundry. (It is assumed that each of the $2n$ socks has precisely one mate.) When the laundry is returned, the socks are drawn out one at a time. Each is matched with its mate, if the mate has previously been drawn. Find a formula for the expected number $E(k)$ of pairs formed after k socks have been drawn.

Solution I by K. Grünbaum and S. Pedersen, Copenhagen V., Denmark. The probability that the ith and jth socks form a matching pair is $1/(2n-1)$. For $1 \leqslant i < j \leqslant k$, let $X_{ij} = 1$ if they match, else $X_{ij} = 0$. Then the number of pairs drawn is $\sum_{1 \leqslant i < j \leqslant k} X_{ij}$, so the expectation $E(k)$ is $E(\sum X_{ij}) = \sum E(X_{ij}) = \binom{k}{2}/(2n-1)$.

Solution II by Mark Bowron, Lynnwood, WA. The probability that pair i is present among the first k socks is $\binom{2n-2}{k-2}\Big/\binom{2n}{k}$. Let $Y_i = 1$ if pair i is present, else $Y_i = 0$. Then the number of pairs present is $\sum_{i=1}^{n} Y_i$, and the expectation is

$$E\left(\sum Y_i\right) = \sum E(Y_i) = n\binom{2n-2}{k-2}\Big/\binom{2n}{k} = \tfrac{1}{2}k(k-1)/(2n-1).$$

Solution III by Richard A. Groeneveld, Iowa State University, Ames. Let X_k be the number of pairs contained in the first k socks. Let $Y_{k+1} = 1$ if the $k+1$st sock matches a previously drawn sock, else $Y_{k+1} = 0$. The conditional probability is $\Pr(Y_{k+1} = 1 | X_k) = (k - 2X_k)/(2n-k)$. The conditional expectation is $E(X_{k+1}|X_k) = E(X_k + Y_{k+1}|X_k) = X_k + (k - 2X_k)/(2n-k)$, which implies $E(k+1) = E(k) + (k - 2E(k))/(2n-k)$. The recurrence, with initial condition $E(1) = 0$, is easily solved to obtain $E(k) = \frac{1}{2}k(k-1)/(2n-1)$.

Editorial comments. Several readers computed the full probability distribution for the number of pairs present among the first k socks and then computed the expectation of that directly. Several readers generalized the problem to r-legged beings, for $r > 2$. More generally, if we have n sets of socks from creatures with $a_1, a_2, \ldots, a_n$ legs respectively, D. E. Knuth remarked that the expected number of complete sets of matching socks after k socks have been drawn at random is

$$\sum_{j=1}^{n} \binom{k}{a_j}\Big/\binom{s}{a_j},$$

where $s = a_1 + a_2 + \cdots + a_n$. Many commented that this problem illustrates the power of the linearity of the expectation over dependent random variables, as used in all solutions above. R. W. van der Waall

(Netherlands) pointed out that for even k the problem (with three solutions) appeared in Dutch in Nieuw Archief voor Wiskunde, Series 3, 15 (1967), 86–87 (problem 123).

Solved by 42 readers and the proposer. Partially solved by two others.

An Exponential Inequality

E 3151 [1986, 401]. *Proposed by Peter Ivady, Institute for Economy and Organization, Budapest, Hungary.*

Let $x \geqslant 0$, $x \neq 1$, $\lambda \geqslant 1$ and $0 \leqslant \beta \leqslant 2$ be real numbers. Prove that

$$\left(\frac{x^{\lambda}-1}{x-1}\right)^{\beta} \leqslant \lambda\left(\frac{x^{\lambda\beta}-1}{x^{\beta}-1}\right).$$

Solution by M. S. Klamkin, University of Alberta. Replacing x by $1/x$ leaves the inequality unchanged, so it suffices to consider only $x > 1$ (it is trivial for $x = 0$). Because $(e^{2\lambda at}-1)/(e^{2at}-1) = (e^{\lambda at}/e^{at}) \cdot (\sinh \lambda at/\sinh at)$, the hyperbolic substitution $x = e^{2t}$ converts the inequality to

$$\lambda\frac{\sinh \lambda\beta t}{\sinh^{\beta}\lambda t} \geqslant \frac{\sinh \beta t}{\sinh^{\beta} t} \tag{1}$$

for $t > 0$, $\lambda \geqslant 1$, and $2 \geqslant \beta \geqslant 0$.

Equation (1) holds with equality for $\lambda = 1$, so it suffices to show that the left side is a nondecreasing function of λ, or equivalently that its logarithmic derivative with respect to λ is non-negative, i.e., $(1/\lambda) + \beta t \coth \lambda\beta t - \beta t \coth \lambda t \geqslant 0$. By multiplying through by $\lambda \sinh \lambda t \cdot \sinh \lambda\beta t$ and using the addition formula for sinh, we transform this inequality into

$$\sinh \lambda t \sinh \lambda\beta t \geqslant \lambda\beta t \sinh \lambda t(\beta - 1). \tag{2}$$

Since sinh is negative for negative arguments, (2) holds for $1 \geqslant \beta \geqslant 0$, and we need only consider $2 \geqslant \beta \geqslant 1$. At $\beta = 2$, (2) reduces to $(\sinh \lambda t)(\sinh 2\lambda t - 2\lambda t) \geqslant 0$, which follows from $\sinh y \geqslant y$ for $y \geqslant 0$. To establish (2) for $2 > \beta \geqslant 1$, it suffices to show that the logarithmic derivative of the left side with respect to β is less than that of the right side. This reduces to showing

$$\frac{\lambda t}{\tanh \lambda\beta t} \leqslant \frac{1}{\beta} + \frac{\lambda t}{\tanh \lambda t(\beta - 1)}.$$

This follows immediately from the fact that tanh is an increasing function.

Also solved by E. Grosswald, V. Pambuccian (Romania), R. E. Shafer, and the proposer.

Partitioning a Collection of Infinite Sets

E 3152 [1986, 401]. *Proposed by Leopoldo Nachbin, University of Rochester, Rochester, NY.*

Let S be a collection of infinite sets. Consider the following partition property (PTP): For every $X \in S$, infinite subsets $Y_X \subset X$ can be assigned such that $Y_{X_1} \cap Y_{X_2} = \varnothing$ if $X_1, X_2 \in S$ are distinct. Prove that

1) S has property (PTP) if it is countable;

2) For every uncountable cardinal number N, there is some S whose power is N, but which fails to have property (PTP).

Solution by Kenneth Schilling, University of Michigan, Flint. (1) Suppose that $S = \{X_i\colon i \in \mathbb{N}\}$, where $\mathbb{N}$ denotes the natural numbers. Let $(p_i\colon i \in \mathbb{N})$ be an enumeration of $\mathbb{N} \times \mathbb{N}$. We now choose elements $y_i \in U_j X_j$ recursively. If $p_t = (m, n)$, let y_t be any element of X_m other than $y_0, y_1, \ldots, y_{t-1}$. Finally, for $m \in \mathbb{N}$, let $Y_{X_m} = \{y_i\colon p_i = (m, n)$ for some $n \in \mathbb{N}\}$. Since p_t runs through all of $\mathbb{N} \times \mathbb{N}$, each set Y_{X_m} is countably infinite, and by construction the sets Y_X are disjoint for distinct $X_i \in S$.

(2) Let S be any uncountable collection of infinite sets, uncountably many of which are subsets of $\mathbb{N}$. Then if sets Y_X, $X \in S$, exist as stated in the problem (even if the condition "Y_X infinite" is weakened to "Y_X nonempty"), then $\cup\{Y_X\colon X \subseteq \mathbb{N}, X \in S\}$ must be uncountable. Since this set is a subset of $\mathbb{N}$, this is impossible.

Editorial comment. Problems like (1) are discussed in great generality by P. Erdös, F. Galvin, and R. Rado in "Transversals and multitransversals", *J. London Math. Soc.* (2) 20 (1979) 387–395. See also W. Sierpiński, *Cardinal and Ordinal Numbers*, p. 459. Assertion (2) is equivalent to a restatement of the theorem on p. 95 of Paul Alexandroff et Paul Urysohn, "Mémoire sur les espaces topologiques compacts," *Verhandelingen Kon. Adad. Wetensch. Amsterdam*, 14 (1929) No. 1.

Also solved by R. E. Bernstein, R. Gilmer, S. Gudder and J. Hagler, Humboldt State Univ. Problem Group, T. Jager, R. Levy, O. P. Lossers (Netherlands), O. Matouš (Czechoslovakia), E. Mendelson, Univ. of Newcastle Problem Solving Class (Australia), V. Pambuccian (Romania), A. K. Wayman and L. Janos, and the referee.

ADVANCED PROBLEMS

For instructions about submitting solutions of Problems, which should be mailed before August 31, 1988, see the inside front cover. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label if an acknowledgment is desired.

6570. *Proposed by L. A. Rubel, University of Illinois at Urbana-Champaign.*

(a) Let (z_n) and (z_n') be sequences in $\mathbb{C}$, neither with a finite limit point. Assume that if a complex number w occurs exactly k times in (z_n) where $k \geqslant 1$, then it occurs exactly $k-1$ times in (z_n'). (Subject to this restriction, we allow finite or even empty sequences.) Show that there exists an entire function f such that the zeros of f are exactly at the points of (z_n) and the zeros of f' are exactly at the points of (z_n'), with the proper multiplicity in each case.

(b) Can one similarly prescribe three sequences $(z_n), (z_n'), (z_n'')$ with corresponding assertions about the zeros of f, f', and f''?

6571. *Proposed by Glenn Ierley, Michigan Technological University, Houghton, MI.*

(a) Let $A(n)$ be the maximum area of a polygon with n sides of lengths $1, 2, \ldots, n$, where $n \geqslant 4$. It is known that the maximum area occurs for a polygon

inscribed in a circle. (Cf. G. Pólya, *Mathematics and Plausible Reasoning*, Volume 1, Princeton, 1954, pp. 174–177.) Let $B(n)$ be the area of a regular polygon with n sides and perimeter $1 + 2 + \cdots + n$. Prove that

$$1 - \frac{A(n)}{B(n)} \sim \frac{\pi^2}{3n^2}. \qquad (n \to \infty).$$

(b) For $1/2 < q < 1$ let $A(q, n)$ be the maximum area of a polygon with n sides of lengths $1, q, q^2, \ldots, q^{n-1}$, respectively, where n is large enough so that $q + q^2 + \cdots + q^{n-1} > 1$. Let $B(q, n)$ be the area of a regular polygon with n sides and perimeter $1 + q + q^2 + \cdots + q^{n-1}$. Prove that $c(q) = \lim_{n \to \infty} A(q, n)/B(q, n)$ exists and find

$$\lim_{q \to 1^-} \frac{1 - c(q)}{(1 - q)^2}.$$

SOLUTIONS OF ADVANCED PROBLEMS

Persistence of a Distribution Function

6522 [1986, 485]. *Proposed by Gunnar Blom, University of Lund and Lund Institute of Technology, Sweden.*

Let $X_1, X_2, \ldots$ be an infinite sequence of independent random variables with the common continuous distribution function F. Let X_N be the first variable that is less than exactly one of all its predecessors $X_1, \ldots, X_{N-1}$. Determine the distribution function of X_N.

Solution by Robert B. Israel, University of British Columbia, Vancouver, BC, Canada. The distribution function of X_N is F. In fact, for any positive integer m this statement is true if "exactly one" is replaced by "exactly m".

For any $k > m$, the probability that X_k is less than exactly m of its predecessors is $1/k$ (since X_k is equally likely to be the first, second, $\ldots$, kth order statistic). Note that this is independent of the ordering of $X_1, \ldots, X_{k-1}$ among themselves. Thus the probability that X_k is the first one less than exactly m of its predecessors is

$$P(X_k = X_N) = \left(1 - \frac{1}{m+1}\right)\left(1 - \frac{1}{m+2}\right) \cdots \left(1 - \frac{1}{k-1}\right)\frac{1}{k} = \frac{m}{k(k-1)}.$$

Since the set of values of a given set of i.i.d. random variables is independent of the order in which they occur, the conditional distribution of X_N given $X_N = X_k$ is the distribution of the $(k - m)$th order statistic for $X_1, \ldots, X_k$. Namely, for any x,

if $P(X_i \le x) = p$, and $q = 1 - p$, then

$$P(X_N \le x | X_N = X_k) = \sum_{j=0}^{m} \binom{k}{j} p^{k-j} q^j$$

so that

$$P(X_N \le x) = \sum_{k=m+1}^{\infty} \frac{m}{k(k-1)} \sum_{j=0}^{m} \binom{k}{j} p^{k-j} q^j.$$

Write this as R_m in order to make the dependence on m explicit. We shall prove by induction on m that $R_m = p$. Note that the series converges absolutely for $0 \le p \le 1$.

First consider the case $m = 1$ (which is the problem as stated). We have

$$R_1 = \sum_{k=2}^{\infty} \frac{1}{k(k-1)} (p^k + kp^{k-1}q)$$

$$= \sum_{k=2}^{\infty} \frac{1}{k(k-1)} (kp^{k-1} - (k-1)p^k)$$

$$= \sum_{k=2}^{\infty} \left(\frac{p^{k-1}}{k-1} - \frac{p^k}{k} \right) = p.$$

Now suppose that $R_{m-1} = p$. We have

$$R_m = \frac{m}{m-1} R_{m-1} - \frac{1}{m-1} \sum_{j=0}^{m-1} \binom{m}{j} p^{m-j} q^j + \sum_{k=m+1}^{\infty} \frac{m}{k(k-1)} \binom{k}{m} p^{k-m} q^m,$$

where the sum over j comprises the terms for $k = m$ that are present in R_{m-1} but not in R_m, and the sum over k comprises the terms for $j = m$ that are present in R_m but not R_{m-1}. Now

$$\sum_{j=0}^{m-1} \binom{m}{j} p^{m-j} q^j = (p+q)^m - q^m = 1 - q^m$$

while

$$\sum_{k=m+1}^{\infty} \frac{m}{k(k-1)} \binom{k}{m} p^{k-m} q^m = q^m \sum_{i=1}^{\infty} \frac{(i+m-2)!}{i!(m-1)!} p^i$$

$$= \frac{q^m}{m-1} (q^{1-m} - 1) = \frac{1}{m-1} (q - q^m).$$

(Here we have used the binomial series

$$\sum_{i=0}^{\infty} \frac{(i+n)!}{i!n!} p^i = (1-p)^{-1-n},$$

which converges for $|p| < 1$; for $p = 1$ the formula is trivial.) Thus

$$R_m = \frac{m}{m-1}p - \frac{1}{m-1}(1 - q^m) + \frac{1}{m-1}(q - q^m) = p$$

as required.

The above generalization of 6522 was also proved by Barthel W. Huff, Eugene Salamin, and Glenn A. Stoops. Both Marcel F. Neuts and the proposer remark that it is sufficient to establish the result for the case in which F is the uniform distribution on $(0, 1)$. The proposer (who also provided a solution for $m = 1$ based on order statistics) used this to provide a noncalculational argument, based on the notion of "records," for the truth of the result. He adds that the result exists in the literature on "records" and is implicit in a paper of Charles M. Goldie and L. C. G. Rogers, The k-record Processes are i.i.d., *Z. für Wahrscheinlichkeitstheorie*, 67 (1984) 197–211.

Neuts added the following remark to his solution. "This result is quite remarkable. It shows that the distribution of the first near-record X_N is the same as that of the underlying random variables. This would be very difficult to infer, for example, from simulation runs. Because of the heavy tail of the distribution of N, the empirical distribution of X_N over many replicated runs may be expected to converge only very slowly to F."

Also solved by Thomas N. Delmer, Ellen Hertz, James M. Meehan, G. S. Rogers, Kenneth Schilling, David G. Weinman, Western Maryland College Problems Group, and Douglas P. Wiens (Canada).

REVIEWS

EDITED BY JOSEPH KONHAUSER
DEPARTMENT OF MATHEMATICS, MACALESTER COLLEGE, ST. PAUL, MN 55015

Combinatorics Theory and Applications. By V. Krishnamurthy. East-West Press PVT LTD, 1985. vi + 483 pp. $84.50.

Applied Combinatorics. By Alan Tucker. John Wiley & Sons, Inc. 1984. v + 447 pp. $31.95

Combinatorics for Computer Science. By S. Gill Williamson. Computer Science Press, 1985. vi + 479 pp.

Constructive Combinatorics. By Dennis Stanton and Dennis White. Springer-Verlag, New York, 1986. x + 183 pp. $19.80

ALBERT NIJENHUIS
Department of Mathematics, University of Pennsylvania, Philadelphia PA 19104

The last few years have seen the appearance of a handful of undergraduate texts in combinatorics. They reflect a renewed interest in what is beyond doubt one of the oldest branches of mathematics. Pressed by questions in various parts of science, many colleges and universities are introducing courses that help biologists analyze DNA chains or trains of nerve impulses, or that help computer scientists set up models for hardware and software designs. Social scientists find combinatorial tools useful to represent relationships between the objects of their study.

Oddly enough, while combinatorics ties in well with probability theory and statistics, its relationship with other branches of "pure" mathematics often seems to be limited to binomial coefficients and an occasional Stirling number. In any case, the interest in teaching combinatorics in many mathematics departments has so far been disappointing.

Still, a strong case can be made for the teaching of combinatorics. First of all, it is a fun subject. Students love it because it is something they really can sink their teeth into. And teachers need not fear: it is an easy field to get into. I recently taught one semester of combinatorics from the Tucker book listed above, and I would call it a success. There are plenty of problems to do, and they are not like the usual calculus problems: find a similar worked-out example and plug in new numbers. Here the logical structure of the subject is so simple that students actually understand what they are doing. After two years of rote calculus, that is a welcome change. This is the first time that they have again a sense of control: concrete constructs and non-trivial relationships.

Let me contrast this with an abstract algebra course I taught. We did groups and rings, after linear algebra. Still, mildly speaking, it was a great disappointment to me

and my colleagues. The difficulty is not in the logical complication (what is simpler than a group?), but in the incredible generality of the group concept. In such circumstances the simplicity of the proofs is of no help: it re-inforces the feeling that the subject is just thin air. To make it more concrete, you have to do lots of examples: permutations, Rubik's cube, and what-have-you. But time does not permit that. So, we deal with only a tiny handful of examples. An aggressive student, after having seen the cyclic group of order p, may look for a non-abelian group of order p^2. Too bad for the time spent; later he learns there aren't any. And chances are he never gets to see a non-abelian group of order p^3, though he does learn theorems about them. He also learns about centers of groups, but how many centers does he get to see? —I could go on.

I believe that combinatorics is not only a field rich in applications in other fields; it is also a good preparation for abstract algebra. At the very least, permutation groups have lots of combinatorial applications. Combinatorics gives you some objects that are worth permuting. And you get to see Burnside's Lemma, and perhaps the counting theorem of Pólya. By contrast, a number of algebra books I checked don't believe in having groups act on sets of real things, and don't even mention the counting theorems. Combinatorics also provides a rich variety of mappings. Some are isomorphisms of structures that students can touch with their own hands. Calculus certainly did not provide any examples of homomorphisms, but combinatorics does. The canonical map from a thin-air group to an even thinner-air factor group is a cruel early example of a structure-preserving map.

The topics that are covered by the numerous books on combinatorics, including the four listed above, differ considerably. This is not just due to the differing aims of the books, but just as much to the absence of a tradition in a subject that is so new to the curriculum. Elementary texts usually cover the following topics, to varying degrees.

Sets and relations: some try to make it a big subject, while others consider it more a matter of setting up decent notation.

Induction should be included, and usually is. This is one of the ideas least understood by immature students. All know vaguely that it takes the place of the "..." in proofs, but only well-designed exercises illustrate its subtleties.

Logic is included in some, but often omitted. It is very hard to produce a body of theory that is truly relevant, but some logical notation is very useful.

Counting is one of the most important topics; in fact so much so that doing just that would make a fairly decent course. It certainly is a very good place to start a course and get the students to work.

There are a few refinements of counting. One is listing the objects you are counting. But usually there are so many that you are forced to develop algorithms, and let a computer do the work.

Another refinement of counting is ranking and unranking. What is the place in a listing that a particular object would have? (Determine it without doing the actual listing!) And conversely, given an integer, what is the object in your list at that address?

Counting techniques include bijections. When you find out, by a nonconstructive method, that two interesting sets have the same number of elements, can you produce an intelligent bijection? If you do, chances are you have learned something worthwhile.

The result of a counting is not always a neat little formula. Sometimes you get a recurrence relation, perhaps similar to the formulas in a Pascal triangle, which can then be further analyzed, either exactly or only asymptotically. Generating functions, usually formal power series, play an essential role and are fun to work with. Inclusion-exclusion is another instructive counting method.

Graph theory is a very large subject. Opinions differ as to whether all of it belongs to combinatorics, or only to the larger area of discrete mathematics. There are so many different approaches, ranging from graph colorings and imbeddings to network flows and sophisticated notions of connectivity. Social scientists, computer scientists, and biologists are likely to ask different questions that have little more in common than the definition of a graph.

Tableaux, coding theory, finite state machines and the analysis of algorithms are just a few more topics that one finds in some of the texts. One can debate whether they are central to the subject, though each is very important at a more advanced level.

Now a few words about the books listed at the beginning, which were used as a springboard for the general comments.

The text by Krishnamurthy is rather fast moving, and at a graduate level. It does start from the beginning, whatever that is, and expects considerable maturity. Much of the space is devoted to collections of exercises. It provides a useful background for anyone teaching combinatorics. It contains some chapters on Schur functions and the characters of the symmetric group.

The book by Stanton and White expects some general background in combinatorics, and deals in depth with a small number of topics: listings, partially ordered sets, bijections, and involutions. Its presentation is lively, and includes some algorithms that actually prove theorems. Its approach is hardly traditional, and brings lots of new ideas. I would recommend that readers have a computer at their side, to test some of their own impulses.

Tucker's book is truly for beginners, at the freshman or sophomore level. It gives very few theorems, but lots of exercises. I did Part II (enumeration) before Part I (graphs, etc.). The level is just a bit below what one would choose for slightly older (3rd or 4th year) students because of the few theorems. No doubt the authors deliberately mentioned matchings but no matching theorems, network flows but not the max-flow-min-cut theorem, etc.

Williamson's book contains a very large number of topics, specifically aimed at computer scientists. Its elaborate table of contents helps a reader in search of a definition or theorem. But if you decide to use this book in a course, you will have to provide additional background. For example, it starts with sets, but does not define cartesian products. It does define relations quite formally, but does not define functions as special kinds of relations. The word "induction" does not appear in the index, nor is it explained. It is used extensively, of course. An early exercise asks for a "careful proof," but at that point the author has not yet given any such proofs, thus keeping the reader guessing as to what that means. And the instructor will also have to supply definitions for "tree" and "rooted tree," since the book only deals with a special case. Similarly, "tiling of a board" and "backtracking" will have to be defined by the teacher. These are just a few examples.

Fortunately, there is an ample choice of books on combinatorics, at virtually all levels. The reader should have no difficulty locating a suitable one.

Applied and Computational Complex Analysis. By Peter Henrici. Wiley-Interscience, 1986. vii 637 pp.

LOUIS AUSLANDER
Department of Mathematics, The Graduate School and University Center, The City University of New York, New York, NY 10036

This book is not a "textbook" in the ordinary sense of the word. The usual textbook is something that is trying to be all things to all people. The late Professor Henrici has written a book that is the distillation of his life's work as a scholar and research mathematician. The book is a lively and lovely piece of work and, like any truly interesting creation, it is full of tensions and contradictions.

The fact that there is a section on digital signal processing, in a book whose last chapter contains a proof of the Bieberbach Conjecture, may suggest that the author has made a rather eclectic choice of topics, but the flow of ideas is very orderly. The presentation of material is rigorous, but the author avoids overburdening the reader by choosing hypotheses for his results that are weak enough for many practical problems, yet strong enough to yield accessible proofs. This is strikingly demonstrated in his discussion of Cauchy integrals in section 14.1 and the results of Calderon discussed in the notes at the end of the section. There are many existential results, but care is taken to ensure effective computation methods. For instance, in Professor Henrici's discussion in § 14.6, entitled "Cauchy Integrals on Straight Line Segments," he arrives at Theorem 14.6a. This is immediately followed by the statement that "Theorem 14.6a does not express what from a numerical point of view may be its most significant aspect." Professor Henrici then reformulates Theorem 14.6a as Algorithm 14.6b, which provides a numerical method for carrying out a desired calculation.

In the Introduction, he states:

> Authors who primarily write for professional mathematicians may cultivate a style where a large number of facts are presented as concisely and economically as possible. However, the present work is not directed exclusively, and perhaps not even primarily, toward such mathematicians. A lifelong career in teaching this kind of reader has convinced me that, however great their appreciation for the logical coherence of the subject, their even greater concern is why they should be interested in it. Thus, time and again, I have allotted valuable space to the task of motivating what is ahead. Moreover, whenever facts or "theorems" are stated—and there are plenty of these —I have endeavored to find formulations that in their essence are intelligible also to readers who did not memorize all the preceding definitions. If I am accused of wordiness and of being, on occasion, repetitive, this is the price I must pay for attempting to reach a larger audience.

The final contradiction is that this book is, in my opinion, admirably suited for educating mathematicians.

An Outline of Set Theory. By James M. Henle. Problem Books in Mathematics, Springer-Verlag, New York, 1986, viii + 145 pp.

C. SMORYŃSKI
Mathematics Institute, Rijksuniversiteit te Utrecht, 3508 TA Utrecht, Netherlands

That there is something wrong with the present teaching of mathematics is indisputable. Controversy arises when one tries to pinpoint exactly what is wrong,

and it intensifies when blame is apportioned or panaceas are offered. However, there are points upon which agreement can be had—particularly if the points are sufficiently blunted and they are not fingers of blame.

What can we agree on? Mathematical education in its broadest form is a failure: the average educated person is unaware that mathematics is still alive, that it is not the case that all mathematical discoveries were made long ago. At the 1986 International Congress of Mathematicians, the winners of the Fields Medals and the Nevanlinna Prize were asked by reporters about the use of the computer in their work, never imagining that the computer is a businessman's rather than a mathematician's tool (even though they correctly perceived what some mathematicians have not, that the computer is mathematical in nature). This ignorance is a matter of miseducation. Partly it is caused by the ignorance of the teachers (high school mathematics teachers generally receive their degrees in education and learn less mathematics than mathematics majors, usually very little modern mathematics), and partly by the textbooks: it has often been remarked how textbooks in the sciences include some history, while college calculus books, generally the only mathematics texts to include any history, only have a few biographical remarks ranging in length from footnotes to half-page paragraphs.

The mathematical training of mathematicians and engineers could also be improved. Twenty years ago a course in the calculus supplied the equivalent of mathematical maturity, the appropriate prerequisite to upper division course work. Since then, the calculus textbooks have been diluted by the elimination of proofs, the labelling as optional of all conceptual material, and the encouragement of passivity on the part of the student by providing ready-made reviews with boxed formulae. Linear algebra is now supposed to provide the bridge between the calculus and higher mathematics, but I see linear algebra texts (as well as texts on differential equations) being produced by the same visionless authors with the same panoply of brain-softening "aids" to students—the cancer is spreading.

Progressing yet farther away from any point of common agreement, I will even say that there is something wrong with education in higher mathematics. The "Definition-Theorem-Proof" style of textbook writing is the most efficient means of transmitting a large quantity of information and it should not be lightly discarded. In a classroom, the dead facts can be brought to life by an informed instructor. But what happens if the instructor himself was textbook trained? He will not be able to bring the artifacts to life, because they are as much artifacts to him as they are to his students. And what will the textbooks such an instructor writes look like?

How do we breathe new life into the teaching of mathematics? There are two proposals crystalizing in the pedagogic vapours. First there is Harold Edwards' exhortation to "Read the Masters" and his guidebooks for those wishing to do so. At the other extreme is the philosophy of letting the students "discover" for themselves, a philosophy inherent in Springer-Verlag's new series Problem Books in Mathematics. Both proposals work, but I am not equally satisfied with their workings. Whereas I can sing lofty paeans to Edwards and his achievements, I can only caution against putting too much faith in problems courses and the readily misapplied Moore method: For, the life these latter breathe into the subject is not the real thing, but an artificial life like Frankenstein's monster and, however much sympathy the monster may evoke, it is yet a teratological creation lacking a soul. James Henle's *An Outline of Set Theory* exemplifies this poverty of spirit.

Consider the actual historical development of set theory. Working on the uniqueness of Fourier series, Georg Cantor obtained his results first for series convergent at all points, then for series convergent at all but finitely many exceptional points, and finally for series with infinitely many nicely distributed exceptional points. The introduction of ordinal numbers was necessary for a definition of "nicely distributed" and Cantor's research changed directions. For this he drew fire from Leopold Kronecker.

The seemingly tiny fact that the cardinality of the real plane is the same as that of the real line is not so tiny when one considers it in its historical context: to Cantor, one-dimensional and two-dimensional space were two distinct objects and, despite his attempt to prove their dissimilarity, he actually proved them to be virtually the same! In a letter to Richard Dedekind about his construction of a one-one correspondence between the unit interval and the unit square, Cantor wrote "I see it, but I don't believe it." Dedekind replied that dimensionality might still be a genuine concept: he didn't think the one-one correspondence could be continuous. This was in 1874; in 1890 Peano showed that the unit square was a continuous image of the unit interval. Perhaps the map could be one-one as well. It was not until 1909 that L. E. J. Brouwer proved the general invariance of dimension.

Set theory explicitly assumed a foundational role around the turn of the century when Gottlob Frege decided to base all of mathematics on it. Cantor's theories of ordinal and cardinal numbers subsumed the arithmetic of the natural numbers, and both Cantor and Dedekind had set-theoretic constructions of the real numbers from there. Unfortunately, Frege's axiomatic system was flawed—as Bertrand Russell noted. Avoiding the paradoxes was something that worried the philosophers, but it had little or no effect on mathematics or the development of set theory. Ernst Zermelo's axiomatization in 1908 accompanied his second proof of the Well-Ordering Theorem from the Axiom of Choice. Zermelo's first proof had met with much criticism and he supplied his new proof with an explicit list of axioms used so that it may be more convincing. About twenty years later, he offered the world the cumulative hierarchy of sets as an intuitive model of these axioms—actually, of an improved set of axioms resulting from some fine tuning by Abraham Fraenkel and Thoralf Skolem.

This minihistory of set theory falls a bit short of completeness, but, inadequate as it is, it does indicate the dynamic life of the subject and the excitement that could be conveyed to a student with an historically based text, or, at least, a scholarly written one. It hints at what the potential mathematicians among the students may experience in their own lifetimes: the unexpected discovery of something interesting, the incredible surprises and complete reversals of one's beliefs or intuitions, the timespan in the solution of problems, the subsumption of one programme by another, and the slow crystalization of concepts. The mention of Kronecker and the criticism of Zermelo's original proof also bring the human element into the story, both in its destructive and constructive aspects—two aspects that are ever at hand.

Now consider the static picture offered by the problems text. Henle has tried to go beyond the "Definition-Theorem-Problem" format by prefacing the work with a $3\frac{1}{2}$ page introduction including history, philosophy, a statement of intent, and discussions of his method and bias. He also sprinkles occasional two or three sentence remarks of a historical or philosophical nature throughout his book. The

picture he paints can be glimpsed from a few quotes from his history:

> Mathematics is a living creature, growing as occasions demand and circumstances permit. Every now and then it must pause to organize and reflect on what it is and where it comes from. This happened in the sixth century B.C. when Euclid thought he had derived most of the mathematical results known at the time from five postulates. By the end of the nineteenth century, it was ready to happen again... In searching for underlying principles, mathematicians were led naturally to sets... [mention of Russell's paradox]... Over the decades following the discovery of such problems, a collection of principles or axioms was formed which appeared (and still appears) to avoid paradoxes. The system is called Zermelo-Fraenkel set theory or ZF after its originators, Ernst Zermelo and Abraham Fraenkel. In addition to occupying a strategic location in mathematics, ZF is studied for itself by a growing number of mathematicians. The father of modern set theory was Georg Cantor.

The story outlined (I have not omitted anything central) is wholly false, but that is hardly the issue—one of the first corollaries of the general floccinaucinihilipilification of scholarship in mathematics is the irrelevance of history. What is at issue is the one-dimensionality and simplicity (simplistic-icity?) of the artificial life created. Real life—healthy life—is rich in its complexity; artificial life—like Rabbi Loeb's golem or Henle's history—is unhealthy in its simplicity.

If one concomitant of the problem solver's perspective is the attitude that history doesn't matter—any story will do (but, the simpler the better)—another is the attitude that the results themselves are not important. Like the fisherman who doesn't eat fish, the mathematical problem solver is only out for the hunt. And, the Great White Hunter leading the safari selects the game suitable for his clients, who will see but a little of the jungle—or, perhaps, only the savannah and none of the jungle. This is exactly what happens in the problems course, as is illustrated by *An Outline of Set Theory*: nothing is taken very far. Dedekind's construction of the reals stops with the definition of $1/r$, thus almost but not quite proving the reals to be a field. The treatment of ordinal exponentiation is given in connexion with the Goodstein-Kirby-Paris Theorem, a recent "combinatorial independence result" closely tied to the Cantor Normal Form for ordinals less than ε_0. Neither ε_0 nor the normal form are mentioned. Cardinal arithmetic is omitted from the discussion of cardinals. Etc.

Other than to suggest that he be raked over the coals, or at least flogged, for the sentence about Euclid, I do not want to attack Prof. Henle for having written an unscholarly book. In this he has been thoroughly professional. I wish but to attack this highly professional perspective of mathematics as mere problem solving. Although some may attribute the lifelessness and shallowness of his book to his dubious goal of running a problem solving course for average students, I think the blame lies entirely in the too narrow perspective. *An Outline of Set Theory* is a vivid illustration of the need for something more: To instill some life into our mathematics texts, we also need some scholarship.

TELEGRAPHIC REVIEWS

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T:	Textbook	P:	Professional Reading	1–4:	Semesters
C:	Computer Software	L:	Undergraduate Library	**:	Special Emphasis
S:	Supplementary Reading	13:	Grade Level	??:	Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

Mathematics Appreciation, T(13: 1). *Mathematics in Daily Life: Making Decisions and Solving Problems.* Joanne Simpson Growney. McGraw-Hill, 1986, xxiii + 453 pp, $33.95. [ISBN: 0-07-025015-4] Covers the use of elementary mathematics, mostly arithmetic, in problem solving and decision making. Does not emphasize specific mathematical topics. Intended for use in college or university "general education" mathematics courses. Could also be of interest to elementary and secondary school teachers. Algebra is a helpful, but not necessary, prerequisite. Very easy to read. Includes many exercises varying in level of difficulty. RH

Education, P, L. *Implementation Handbook for the Comprehensive Mathematics Program.* Manford Byrd, Jr. (Board of Education of the City of Chicago, 1819 W. Pershing Rd., Chicago, IL 60609), 1987, (P). *Kindergarten-Grade 3,* viii + 154 pp; *Grades 4-6,* viii + 191 pp; *Grades 7-8,* viii + 138 pp. A detailed set of objectives and examples for each topic to be covered in each reporting period in grades K-8 in the Chicago public schools. Includes creative calculator examples, as well as geometry, data analysis, measurement, and applications in each grade. LAS

History, P, L.** *Reminiscences About a Great Physicist: Paul Adrien Maurice Dirac.* Ed: Behram N. Kursunoglu, Eugene P. Wigner. Cambridge U Pr, 1987, xviii + 297 pp, $49.50. [ISBN: 0-521-34013-6] Two dozen reflections by close friends of Dirac on both scientific and personal associations. Contributors include his widow Margit Dirac, her brother Eugene Wigner, P.A.M. Dirac himself (on the inadequacies of quantum field theory), and many world-famous physicists such as Harish-Chandra, Fred Hoyle, and Abdus Salam. LAS

History, S(17), P*. *The Historical Development of Quantum Theory, Volume 5: Erwin Schrödinger and the Rise of Wave Mechanics, Part 2: The Creation of Wave Mechanics; Early Response and Applications 1925-1926.* Jagdish Mehra, Helmut Rechenberg. Springer-Verlag, 1987, ix + 613 pp, $79.95. [ISBN: 0-387-96377-4] Wave mechanics began when Schrödinger showed in his fundamental papers of 1926 that the quantum states of the hydrogen atom could be represented as an eigenvalue problem for an appropriate operator. This volume reconstructs the scientific attitude of this revolutionary period. Schrödinger's own thoughts and motivation have always been obscure due to his lack of correspondence. The authors attempt to address this by presenting in detail the contributions of de Broglie, Einstein, and others along with Schrödinger's own work. (*Part 1,* TR, January 1988.) MR

Foundations, T(14: 1), S**, L*.** *Bridge to Abstract Mathematics: Mathematical Proof and Structures.* Ronald P. Morash. Math. Ser. Random House, 1987, xiii + 395 pp, $24. [ISBN: 0-394-35429-X] A text for a course which bridges the gap between beginning courses like calculus and more advanced courses in which students deal seriously with mathematical proofs. Topics include sets, logic, applications of logic, methods of mathematical proof, relations, functions, and the construction of number systems. Readable with lots of exercises of varying difficulty and solutions to selected exercises. CEC

Foundations, P, L. *Particles and Paradoxes: The Limits of Quantum Logic.* Peter Gibbins. Cambridge U Pr, 1987, xi + 181 pp, $34.50; $11.95 (P). [ISBN: 0-521-33498-5; 0-521-33691-0] An exposition and exposé of attempts to intepret the paradoxes of quantum mechanics (e.g., the twin slit experiment) via quantum logic. The author holds that quantum logic—indeed, any logic—is inadequate to the task of providing a foundation for quantum mechanics. "We are left with a mystery" of "just how odd the physical world must be." LAS

Foundations, T(17-18), S, P, L. *Varieties of Con-*

structive Mathematics. Douglas Bridges, Fred Richman. London Math. Soc. Lect. Note Ser., V. 97. Cambridge U Pr, 1987, x + 149 pp, $19.95 (P). [ISBN: 0-521-31802-5] An introduction to the spirit and practice of modern constructive mathematics, requiring a minimal background in philosophy and formal logic. The book introduces three varieties of constructive mathematics: Bishop's constructive mathematics, Brouwer's intuitionistic mathematics, and the constructive recursive mathematics of the Russian school of Markov. LCL

Combinatorics, T(17-18), P, L. *Combinatorial Geometries.* Ed: Neil White. Encycl. of Math. & Its Applic., V. 29. Cambridge U Pr, 1987, xii + 212 pp, $39.50. [ISBN: 0-521-33339-3] This is volume two in a three-volume series on matroid theory. Topics covered include coordinatizations, matching theory, transversed and simplicial matroids, the Möbius function and combinatorial optimization. Includes exercises. LC

Number Theory, T*(16-18: 1), S, P, L*. *A Course in Number Theory and Cryptography.* Neal Koblitz. Grad. Texts in Math., V. 114. Springer-Verlag, 1987, 208 pp, $34. [ISBN: 0-387-96576-9] A good introduction to the elementary number theory and algebra (e.g., finite fields) underlying public-key cryptosystems, including an exposition of the elliptic curve approach. Unfortunately, the cheap fuzzy printing is unpleasant to look at. Springer shouldn't let this happen in the GTM series. BC

Number Theory, T(14: 1), S, L*.** *Elementary Number Theory.* Charles Vanden Eynden. Math. Ser. Random House, 1987, xii + 266 pp, $27. [ISBN: 0-394-35359-5] A well-written introduction to the subject. Extensive exercise sets include many routine problems as well as a good collection of theoretical problems. Lots of historical material. Advanced topics include public key cryptography, quadratic residues, continued fractions, some diophantine equations, and Pell's equation. CEC

Linear Algebra, T(15-17: 1). *Applied Linear Algebra, Third Edition.* Ben Noble, James W. Daniel. Prentice-Hall, 1988, xvi + 521 pp. [ISBN: 0-13-041260-0] Major changes in this edition include greater emphasis on triangular forms, over 500 new exercises, and additional material on some topics previously discussed only in exercises (e.g., the Cayley-Hamilton and Perron-Frobenius theorems). Some examples and problems using computer software for numerical linear algebra have also been added. (*First Edition,* TR, October 1969; *Second Edition,* TR, November 1978.) AO

Linear Algebra, P. *Invariant Theory and Superalgebras.* Frank D. Grosshans, Gian-Carlo Rota, Joel A. Stein. CBMS Reg. Conf. Ser. in Math., No. 69. AMS, 1987, xxi + 80 pp, $16 (P). [ISBN: 0-8218-0719-6] A superalgebra is one part symmetric, one part exterior, and one part "divided powers." The authors prove an extension of the "standard basis theorem," and apply it to compute invariants of symmetric and skew-symmetric tensors. Includes a very clear and fairly long synopsis, and many examples. BC

Linear Algebra, T?(13: 1). *A First Course in Linear Algebra with Concurrent Examples.* A.G. Hamilton. Cambridge U Pr, 1987, viii + 148 pp, $39.50; $12.95 (P). [ISBN: 0-521-32516-1; 0-521-31041-5] Not enough material for a standard course. Covers solutions to linear equations, linear dependence, determinants through the 3x3 case, and some three-dimensional geometry. Answers to most exercises in back. GG

Linear Algebra, T*(14: 1, 2), L. *Linear Algebra.* John B. Fraleigh, Raymond A. Beauregard. Addison-Wesley, 1987, xv + 519 pp, $35.95. [ISBN: 0-201-15459-5] Includes standard topics: linear systems, vector spaces and linear transformations, determinants, eigenvalues, and orthogonality, all with emphasis on R^n. No Jordan forms. Applications include computer solutions of linear systems and eigenvalues, least squares, and linear programming. Also has chapter on calculus applications. Has answers to odd exercises; software to accompany text available. Good for course with applied flavor. GG

Linear Algebra, T(16-17: 1), S*, L. *Nonnegative Matrices and Applicable Topics in Linear Algebra.* Alexander Graham. Math. & Its Applic. Halsted Pr, 1987, 264 pp, $89.95. [ISBN: 0-470-20855-4] Introduction for general readers, includes background material (including aspects of graph theory; unitary, Hermitian, and normal matrices; positive definite matrices), followed by the main theme of the book and applications. Important concepts amply illustrated with worked examples and problems (with solutions). LCL

Group Theory, P. *Lecture Notes in Mathematics-1281: Group Theory.* Ed: O.H. Kegel, F. Menegazzo, G. Zacher. Springer-Verlag, 1987, vii + 179 pp, $16.30 (P). [ISBN: 0-387-18399-X] A collection of 19 short papers which were presented at the International Conference on Group Theory held at Brixen/Bressanone, Italy, May 25-31, 1986. The papers consider quite a range of topics; many involve nilpotent groups and several concern locally finite groups. LW

Group Theory, S(18), P. *Unitary Representations of Reductive Lie Groups.* David A. Vogan, Jr. Annals of Math. Stud., No. 118. Princeton U Pr, 1987, x + 308 pp, $60; $19.50 (P). [ISBN: 0-691-08481-5; 0-691-08482-3] General survey based loosely on January 1986 Hermann Weyl Lectures given at the Institute for Advanced Study. Introduction gives excellent perspective. Some proofs omitted or sketched. GG

Group Theory, P. *Essays in Group Theory.* Ed: S.M. Gersten. Math. Sci. Res. Inst., V. 8. Springer-Verlag, 1987, 342 pp, $32. [ISBN: 0-387-96618-8] A collection of five papers based on a seminar presented in 1985 at MSRI, Berkeley. Baumslag and Shalen examine presentations by the classical method of representations; Stalling and Gersten deal with applications of diagrams to group theory, and Gromov

and Shalen with hyperbolic groups and actions on R-trees. The papers are all well-written and free of the typical terseness of papers appearing in journals. LW

Algebra, T(18: 1, 2), P. *Commutator Theory for Congruence Modular Varieties.* Ralph Freese, Ralph McKenzie. London Math. Soc. Lect. Note Ser., V. 125. Cambridge U Pr, 1987, 227 pp, $27.95 (P). [ISBN: 0-521-34832-3] A development of a theory of commutators in the setting of congruence modular varieties. This theory generalizes the notion of the commutator of a group. Includes many exercises to solutions and extensive bibliographical notes. SG

Algebra, P. *Noetherian Rings and Their Applications.* Ed: Lance W. Small. Math. Surv. & Mono., No. 24. AMS, 1987, ix + 118 pp, $38. [ISBN: 0-8218-1525-3] A collection of six papers (at just over $6 apiece) on such topics as the Goldie rank of a module, semisimple Lie algebras, and Noetherian group rings. SG

Algebra, P. *Lecture Notes in Mathematics-1271: Algebraic Groups, Utrecht 1986.* Ed: A.M. Cohen, *et al.* Springer-Verlag, 1987, viii + 284 pp, $24.30 (P). [ISBN: 0-387-18234-9] Fourteen papers from the April 1986 symposium honoring the 350th anniversary of the University of Utrecht and T.A. Springer's 60th birthday. GG

Algebra, T(14-15: 1). *Rings and Factorization.* David Sharpe. Cambridge U Pr, 1987, ix + 111 pp, $14.95 (P); $34.50. [ISBN: 0-521-33718-6; 0-521-33072-6] A very readable and enjoyable introduction to the concepts of rings, fields, prime elements, and unique factorization. Assumes no background in abstract algebra. Includes examples of concrete applications, such as factoring polynomials and Fermat's two-squares theorem. Contains many exercises along with hints and solutions. A fun and smooth introduction to some abstract mathematical ideas. RH

Algebra, P. *Structures Paragraduées (Groupes, Anneaux, Modules).* Marc Krasner, Mirjana Vuković. Papers in Pure & Appl. Math., No. 77. Queen's U, 1987, 163 pp, (P).

Algebra, T*(16: 1, 2), S, L*. *Introduction to Abstract Algebra.* Elbert A. Walker. Math. Ser. Random House, 1987, viii + 355 pp, $30. [ISBN: 0-394-35611-X] A sophisticated introduction to algebra for undergraduates. Chapter titles include sets, groups, vector spaces, rings and modules, linear transformations, fields, and topics from both group and ring theory. Includes excellent collections of exercises, but few of them are routine. Plenty of material for a year-long course. CEC

Algebra, P. *Lecture Notes in Mathematics-1280: Jordan Triple Systems by the Grid Approach.* Erhard Neher. Springer-Verlag, 1987, xii + 193 pp, $20 (P). [ISBN: 0-387-18362-0]

Real Analysis, P. *Regular Variation, Extensions and Tauberian Theorems.* J.L. Geluk, L. de Haan. CWI Tract, V. 40. Math Centrum, 1987, 132 pp, Dfl. 20.30 (P). [ISBN: 90-6196-324-9] A self-contained introduction to the theory of regular variation and its main extensions. It is shown that regularly varying functions are a natural setting for Tauberian theorems of the Laplace type. Also includes some results for general kernel transforms. Includes references. CEC

Complex Analysis, S(18), P. *Lectures on Counterexamples in Several Complex Variables.* John Erik Fornæss, Berit Stensønes. Math. Notes 33. Princeton U Pr, 1987, 247 pp, $22.50 (P). [ISBN: 0-691-08456-4] Lecture notes from a graduate course at Princeton. Begins with a brief introduction to basics of several complex variables. Topics include counterexamples to smoothing of plurisubharmonic functions, CR-manifolds, Stein neighborhood basis, peak sets, inner functions, Runge exhaustion. Presented in a clear, but somewhat informal style. No problems. BH

Complex Analysis, P. *Lecture Notes in Mathematics-1275, 1276, 1277: Complex Analysis.* Ed: Carlos A. Berenstein. Springer-Verlag, 1987, (P). *I,* xv + 331 pp, $28.60, [ISBN: 0-387-18356-6]; *II,* ix + 320 pp, $28.60, [ISBN: 0-387-18357-4]; *III,* x + 350 pp, $32.90. [ISBN: 0-387-18355-8] Proceedings of the sixteenth Special Year at the University of Maryland from July 1985 to December 1986. A collection of 54 papers on topics from complex analysis, including both survey articles and new results. RH

Complex Analysis, T*(16-17: 2, 3), L. *Complex Functions: An Algebraic and Geometric Viewpoint.* Gareth A. Jones, David Singerman. Cambridge U Pr, 1987, xiv + 342 pp, $59.50; $16.95 (P). [ISBN: 0-521-30893-3; 0-521-31366-X] Presents "main ideas about complex functions and Riemann surfaces" assuming only fundamentals of abstract algebra, real and complex analysis, and topology. Topics are Riemann sphere, Möbius transformations, elliptic functions, analytic continuation and Riemann surfaces, hyperbolic geometry, and the modular group. Well motivated with good selection of problems at end of each chapter. BH

Differential Equations, P. *Perturbation Methods, Bifurcation Theory and Computer Algebra.* Richard H. Rand, Dieter Armbruster. Appl. Math. Sci., V. 65. Springer-Verlag, 1987, ix + 243 pp, $29.80 (P). [ISBN: 0-387-96589-0] Contains MACSYMA programs for several popular perturbation methods as well as methods for studying bifurcations. A brief introduction to MACSYMA is provided as an appendix. AO

Differential Equations, P. *Ordinary and Partial Differential Equations.* Ed: B.D. Sleeman, R.J. Jarvis. Longman Scientific & Technical (US Distr: Wiley), 1987, 216 pp, $48.95 (P). [ISBN: 0-470-20839-2] Proceedings of the ninth Dundee Conference, 1986, with fourteen papers by seventeen authors; topics include coherent excitations, one- and two-parametric bifurcation diagrams, inverse scattering, limit cycle configurations of quadratic systems, concentration effects in Euler equations, homoclinic orbits and chaos in delay equations, the discrete Nagumo equation, radiation of sound by a moving sphere, fisheries stock assessment, Cahn-Hilliard

model of phase transition, second-order elliptic equations, biHamiltonian systems, applications of differential equations in biology, and fibered structures in optimal design. RSF

Differential Equations, T*(16-17: 1, 2). *Applied Mathematics: A Contemporary Approach.* J. David Logan. Wiley, 1987, xviii + 572 pp, $44.95. [ISBN: 0-471-85083-7] A survey text covering most of the traditional topics and a few nontraditional ones as well: dimensional analysis and scaling, perturbation methods, calculus of variations, partial differential equations of mathematical physics, stability and bifurcation, similarity methods, and numerical techniques for partial differential equations. AO

Partial Differential Equations, P. *Contributions to Nonlinear Partial Differential Equations, Volume II.* Ed: J.I. Díaz, P.L. Lions. Res. Notes in Math. Ser., V. 155. Longman Scientific & Technical (US Distr: Wiley), 1987, 308 pp, $57.95 (P). [ISBN: 0-470-20810-4] The proceedings of the Second Franco-Spanish Colloquium on nonlinear partial differential equations held in Paris in December 1985. CEC

Partial Differential Equations, P. *Lecture Notes in Mathematics-1248: Nonlinear Semigroups, Partial Differential Equations and Attractors.* Ed: T.L. Gill, W.W. Zachary. Springer-Verlag, 1987, ix + 185 pp, $18 (P). [ISBN: 0-387-17741-8] This volume constitutes the proceedings of the symposium on nonlinear semigroups, partial differential equations, and attractors held at Howard University in Washington, D.C. on August 5-8, 1985. CEC

Partial Differential Equations, P. *Lecture Notes in Mathematics-1241: Singularities in Linear Wave Propagation.* Lars Gårding. Springer-Verlag, 1987, 125 pp, $13.10 (P). [ISBN: 0-387-18001-X] The aim of these lectures is to present the use of microlocal theory in the analysis of singularities in linear wave propagation. LCL

Partial Differential Equations, P. *Hyperbolic Equations.* Ed: F. Colombini, M.K.V. Murthy. Res. Notes in Math. Ser., V. 158. Longman Scientific & Technical (US Distr: Wiley), 1987, 286 pp, $51.95 (P). [ISBN: 0-470-20869-4] Proceedings of the conference on Hyperbolic Equations and Related Topics, University of Padova, 1985. LC

Partial Differential Equations, S(18), P. *Viscosity Solutions and Optimal Control.* Robert J. Elliott. Res. Notes in Math. Ser., V. 165. Longman Scientific & Technical (US Distr: Wiley), 1987, 95 pp, $39.95 (P). [ISBN: 0-470-20918-6] Discusses the concept of a viscosity solution and its relationship to a cost/value function of an optimal control problem or differential game. A viscosity function is any solution to the Hamilton-Jacobi-Bellman differential equation. Topics covered include dynamic programming and differential games with alternate, overlapping play. MR

Partial Differential Equations, P. *Wave Motion: Theory, Modelling, and Computation.* Ed: Alexandre J. Chorin, Andrew J. Majda. Math. Sci. Res. Instit., V. 7. Springer-Verlag, 1987, 336 pp, $32. [ISBN: 0-387-96594-7] Twelve papers from the proceedings of a conference given at the Mathematical Sciences Research Institute, Berkeley, California, June 1986, in honor of the 60th birthday of Peter Lax. LCL

Partial Differential Equations, P. *The Homotopy Index and Partial Differential Equations.* Krzysztof P. Rybakowski. Universitext. Springer-Verlag, 1987, xii + 208 pp, $39.50 (P). [ISBN: 0-387-18067-2] Homotopy index theory originated on compact spaces. As usual, the non-compact case is problematic. The author gives a unified approach, with applications to partial differential equations and other topics (e.g., Morse theory). BC

Numerical Analysis, P. *Lecture Notes in Engineering-27: The Best Approximation Method: An Introduction.* Th. V. Hromadka II, Ch.-Ch. Yen, G.F. Pinder. Springer-Verlag, 1987, xiii + 168 pp, $26.70 (P). [ISBN: 0-387-17572-5] A generalized Fourier expansion method for solving linear operator equations. Mathematical background in metric spaces and Lebesgue integration. Application to engineering problems including computer codes. RWN

Numerical Analysis, P. *Algorithms for Approximation.* Ed: J.C. Mason, M.G. Cox. Inst. of Math. & Its Applic. Conf. Ser., V. 10. Clarendon Pr, 1987, xvi + 694 pp, $125. [ISBN: 0-19-853612-7] 41 contributions at an international conference held July 15-19, 1985 at the Royal Military College of Science in Shrivenham, England. They concern algorithm development in spline approximation and smoothing, spline interpolation and shape preservation, multivariate interpolation, least-square methods, rational approximation, complex and nonlinear approximation, computer-aided design and blending; applications in numerical analysis, partial differential equations, other disciplines; software. DFA

Operator Theory, S(18), P. *Summing and Nuclear Norms in Banach Space Theory.* G.J.O. Jameson. Math. Soc. Stud. Texts, V. 8. Cambridge U Pr, 1987, xi + 174 pp, $39.50; $13.95 (P). [ISBN: 0-521-34134-5; 0-521-34937-0] Assumes basic knowledge of Banach and Hilbert space theory. Main topics are Pietsch's theorem on p-summing operators and Grothendieck's inequality and their applications. Exercises and numerous examples scattered throughout. BH

Functional Analysis, T(15: 1). *Introduction to the Analysis of Metric Spaces.* J.R. Giles. Australian Math. Soc. Lect. Ser., V. 3. Cambridge U Pr, 1987, xiv + 257 pp, $49.50; $16.95 (P). [ISBN: 0-521-35051-4; 0-521-35928-7] Follows educational style of Dieudonné used in British universities: detailed study of analysis of metric spaces before introducing topological spaces. Normed linear spaces treated as special subfamily of metric spaces. Very helpful geometrically-suggestive diagrams. Numerous good exercises at the end of each section. Connectedness, Axiom of Choice not included. MS

Functional Analysis, P. *Lecture Notes in Mathematics-1267: Geometrical Aspects of Functional*

Analysis. Ed: J. Lindenstrauss, V.D. Milman. Springer-Verlag, 1987, 212 pp, $20 (P). [ISBN: 0-387-18103-2] Proceedings of the Israel Seminar on geometric aspects of functional analysis held between October 1985 and June 1986. Central topic is convex sets in R^n and infinite-dimensional spaces. BH

Analysis, P. *Lecture Notes in Mathematics-1257: On the C^*-Algebras of Foliations in the Plane.* Xiaolu Wang. Springer-Verlag, 1987, 165 pp, $16.30 (P). [ISBN: 0-387-17903-8] A study of C^*-algebras of foliated open two manifolds, linking topology, graph theory, geometry, and analysis in an important new area of research. LCL

Analysis, P. *Lecture Notes in Mathematics-1265: Asymptotics for Orthogonal Polynomials.* Walter Van Assche. Springer-Verlag, 1987, vi + 201 pp, $20 (P). [ISBN: 0-387-18023-0] A monograph which concentrates on the asymptotic theory of general orthogonal polynomials on the real line. Includes a substantial list of references. CEC

Analysis, P. *A General Theory of Integration in Function Spaces, Including Wiener and Feynman Integration.* P. Muldowney. Res. Notes in Math. Ser., V. 153. Longman Scientific & Technical (US Distr: Wiley), 1986, 115 pp, $34.95 (P). [ISBN: 0-470-20736-1] Begins with a brief description of Henstock's general theory of integration (a simple, but powerful generalization of Riemann integration), then develops theory of function space integration in more detail and applies this theory to statistical mechanics (Wiener integral) and quantum mechanics (Feynman integral), thus giving a unified treatment of these two integrals. BH

Analysis, S(18), P, L. *Moments in Mathematics.* Ed: Henry J. Landau. Proc. of Symp. in Appl. Math., V. 37. AMS, 1987, xi + 154 pp, $30. [ISBN: 0-8218-0114-7] Applications and ramifications of the classical problem of moments (when does a sequence of numbers represent the successive moments of a non-negative measure?) to geometry, Hilbert space, signal processing, semigroups, probability and statistics. Papers from the January 1987 AMS short course in San Antonio. LAS

Analysis, P. *Value Distribution Theory for Meromorphic Maps.* Wilhelm Stoll. Aspects of Math., V. E7. Friedr. Vieweg & Sohn, 1985, xi + 347 pp, $28 (P). [ISBN: 3-528-08906-7] Generalizes standard Nevanlinna theory to the following setting: $f : M \to P_n$ is a meromorphic map, M is a parabolic manifold of dimension m, and P_n is n-dimensional projective space. Extensive bibliography. BH

Analysis, T(16-17). *Finite-Dimensional Spaces: Algebra, Geometry, and Analysis, Volume I.* Walter Noll. Mech.: Analy., V. 10. Martinus Nijhoff (US Distr: Kluwer Academic), 1987, xvi + 393 pp, $97.50. [ISBN: 90-247-3581-5] Based on a course originally entitled "Tensor Analysis" and then renamed "Multidimensional Algebra, Geometry, and Analysis," the text attempts to unify these topics and "to present mathematics as a whole." The prerequisites for a student are courses in abstract algebra, real analysis, and linear algebra. The topics covered include: dual, flat (affine), and inner-product spaces; basic point-set topology; basic differential calculus; spectral theory (concentrating on finite-dimensional spaces); and the structure of linear transformations. Each of the ten chapters is followed by a problem set. LW

Algebraic Geometry, S(18), P. *Géométrie algébrique réelle.* J. Bochnak, M. Coste, M-F. Roy. Ser. of Mod. Surv. in Math., Band 12. Springer-Verlag, 1987, x + 373 pp, $118.30. [ISBN: 0-387-16951-2] If you want to make algebraic geometry *really* complex, do it over R, not C. BC

Differential Geometry, T(18: 1), S*, P*. *Teichmüller Theory and Quadratic Differentials.* Frederick P. Gardiner. Wiley, 1987, xvii + 236 pp, $46.95. [ISBN: 0-471-84539-6] Treats the moduli space of inequivalent Riemann surface structures on a given topological surface. Full proofs and exercises make the book self-contained for a reader with knowledge of a few topics beyond the basics of complex variables, differential geometry, and algebraic topology. GG

Geometry, S, L.** *A Budget of Trisections.* Underwood Dudley. Springer-Verlag, 1987, xv + 169 pp, $29.80. [ISBN: 0-387-96568-8] A witty, compassionate, and richly personal account of the countless amateur mathematical Don Quixotes who pursue their dream of trisecting angles. Echoing De Morgan's 1872 *A Budget of Paradoxes,* Dudley describes the trisectors and their trisections, showing that despite their ignorance of advanced mathematics, trisectors are often solid citizens who through clever persistence produce some very good approximations. LAS

Geometry, T(17-18: 2, 3), S, P. *Algorithms in Combinatorial Geometry.* Herbert Edelsbrunner. EATCS Mono. on Theor. Comput. Sci., V. 10. Springer-Verlag, 1987, xv + 423 pp, $49. [ISBN: 0-387-13722-X] Despite its title, this book offers an extensive introduction to combinatorial geometry, generally, and to its correlative field, computational geometry, before treating geometric algorithms and their applications. Exercises, research problems, and up-to-date bibliography. SS

Geometry, P. *Intuitive Geometry.* Ed: K. Böröczky, G. Fejes Tóth. Colloquia Mathematica Societatis Janos Bolyai, V. 48. Elsevier Science, 1987, 708 pp, Dfl. 295. [ISBN: 0-444-879-33-1] Assortment of research papers on convexity, discrete and combinatorial geometries, and relations between geometry and other fields: biology, architecture, art, and psychology. SS

Algebraic Topology, P. *Lecture Notes in Mathematics-1274: Equivariant K-Theory and Freeness of Group Actions on C^*-Algebras.* N. Christopher Phillips. Springer-Verlag, 1987, viii + 371 pp, $32.90 (P). [ISBN: 0-387-18277-2] Given an action of a compact Lie group on a compact Hausdorff space, an examination of the equivariant K-theory reveals whether this group action is free. The category of

compact Hausdorff spaces is (contravariantly) equivalent to the category of commutative unital C^*-algebras, so it becomes natural to try to extend the above result to general C^*-algebras. There being no completely satisfactory notion of freeness of an action on a C^*-algebra, the author herein defines and studies conditions on the equivariant K-theory of a C^*-algebra, and in case the algebra is commutative, is able to determine through the equivariant K-theory whether the underlying space is free. LW

Differential Topology, P. *Lecture Notes in Mathematics-1264: Rational Homotopy Theory.* Wu Wen-tsün. Springer-Verlag, 1987, viii + 219 pp, $20 (P). [ISBN: 0-387-13611-8] Clarifying and extending the works of D. Sullivan, this book studies the rational homotopy type (minimal model) of a differential graded algebra of differential forms on a simplicial complex K. In particular, the book shows that the minimal model is homotopically invariant, and contains complete information on the cohomology and homotopy ring of K. The author takes a constructional approach, and is thus able to explicitly determine the minimal model of a fiber space as well as that of a fiber-square-constructed space. LW

Topology, T*(17-18), S*, P, L. *Classical Tessellations and Three-Manifolds.* José Maria Montesinos. Universitext. Springer-Verlag, 1987, xvii + 230 pp, $35 (P). [ISBN: 0-387-15291-1] Written to provide graduate students with a source of geometrical intuition in low-dimensional topology, this fascinating monograph illuminates three of Thurston's eight geometries by exploring manifolds of tessellations in Euclidean, spherical, and hyperbolic space. Profusely illustrated with mineral crystal patterns, ornamental motifs, and mathematical sketching. LAS

Dynamical Systems, T(16-17), P, L. *Chaotic Vibrations: An Introduction for Applied Scientists and Engineers.* Francis C. Moon. Wiley, 1987, xv + 309 pp, $39.95. [ISBN: 0-471-85685-1] An applied view of chaotic phenomena, richly illustrated with physical systems, mathematical models, and standard examples (Henon maps, Lorenz attractors, Poincaré maps, Lyapunov exponents). Little direct mathematical theory, but a great deal of illustrative applications that help to develop intuition. LAS

Dynamical Systems, P. *Attraktoreingrenzung für nichtlineare Systeme.* Gennadij A. Leonov, Volker Reitmann. Teubner-Texte zur Math., V. 97. BG Teubner, 1987, 196 pp, 20M (P). [ISBN: 3-322-00427-9]

Dynamical Systems, P. *Autowave Processes in Kinetic Systems: Spatial and Temporal Self-Organization in Physics, Chemistry, Biology, and Medicine.* V.A. Vasiliev, *et al.* Math. & Its Applic. D Reidel (US Distr: Kluwer Academic), 1987, 262 pp, $69. [ISBN: 90-277-2379-6] From Chapter 1: "An autowave process is a self-sustained wave process in an active nonlinear medium maintaining its characteristics at a constant level at the expense of an energy source distributed in the medium." The key word is "nonlinear." Examples include propagation of phase-transition fronts, Belousov-Zhabotinskii reactions, and nerve cell activity. BC

Probability, S, L**.** *Fifty Challenging Problems in Probability with Solutions.* Frederick Mosteller. Dover, 1987, viii + 88 pp, $3.95 (P). [ISBN: 0-486-65355-2] Unabridged republication of the work first published in 1965. Stimulating and challenging problems with leisurely, instructive, conversational solutions that demonstrate mathematical thinking at its best. LCL

Probability, P*. *Dependence in Probability and Statistics: A Survey of Recent Results.* Ed: Ernst Eberlein, Murad S. Taqqu. Prog. in Prob. & Stat., V. 11. Birkhauser Boston, 1986, xi + 473 pp, $39.50. [ISBN: 0-8176-3323-5] Twenty-three papers written by participants in an international conference held in Oberwolfach in April 1985, concerned with limiting results assuming various dependence structures. "These papers tie together known results, describe the underlying ideas, summarize the state-of-the-art, and state some open problems." RSK

Stochastic Processes, P*. *Stochastic Geometry and Its Applications.* D. Stoyan, W.S. Kendall, J. Mecke. Prob. & Math. Stat. Wiley, 1987, 345 pp, $49.95. [ISBN: 0-471-90519-4] Revised translation of Stoyan and Mecke's 1983 book *Stochastische Geometrie.* "Topics covered include the basic theories of point processes, random sets, fibre processes, tessellations, stereology and the statistical theory of shape." Extensive bibliography of this relatively new field. RSK

Stochastic Processes, S(17-18), P. *Radically Elementary Probability Theory.* Edward Nelson. Annals of Math. Stud., No. 117. Princeton U Pr, 1987, ix + 97 pp, $40; $15 (P). [ISBN: 0-691-08473-4; 0-691-08474-2] Develops the theory of probability and stochastic processes using nonstandard analysis, leaving the results in nonstandard form. More advanced appendix shows how the conventional theory of stochastic processes can be derived from nonstandard theory. Note hardcover price. RSK

Stochastic Processes, T(18: 1, 2). *Applied Probability and Queues.* Søren Asmussen. Prob. & Math. Stat. Wiley, 1987, x + 318 pp, $67.95. [ISBN: 0-471-91173-9] Divided into three parts: Part A deals with simple Markovian models, particularly as they relate to queueing theory; Part B presents basic mathematical tools—renewal theory, regenerative processes, and random walks; Part C investigates special models and methods, primarily dealing "with a more narrow class of problems associated with general distributions of interarrival times and service times." RSK

Elementary Statistics, T*(13: 1), S. *Introduction to Biostatistics, Second Edition.* Robert R. Sokal, F. James Rohlf. WH Freeman, 1987, xii + 363 pp, $37.95. [ISBN: 0-716-71805-7] Modest revision of the authors' 1973 *First Edition* (TR, March 1974), again based on their more inclusive book *Biometry (Second Edition,* TR, May 1982). Changes in-

clude some new descriptive procedures, more nonparametrics, the Bonferroni method of multiple comparisons, and the use of the G statistic rather than the more common chi-square test to analyze frequency data. RSK

Elementary Statistics, T(14-16), S, L. *Elements of Statistics for the Life and Social Sciences.* Braxton M. Alfred. Texts in Stat. Springer-Verlag, 1987, xiii + 190 pp, $36. [ISBN: 0-387-96500-9] Expansive, self-contained introduction to statistical ideas needed for scientific work in anthropology (and presumably other life and social sciences as well). Case studies used to illustrate logical argument, deductive thinking, prediction, and hypothesis testing. No exercises. LCL

Elementary Statistics, T(14-15: 1, 2). *Basic Statistical Methods for Engineers and Scientists, Third Edition.* John B. Kennedy, Adam M. Neville. Harper & Row, 1986, xx + 613 pp. [ISBN: 0-06-043633-6] Revision of the authors' 1976 *Second Edition.* Includes new material on probability, statistical inference, and goodness of fit. Emphasizing practical considerations, it also includes chapters on rejection of outliers, distributions of extremes, tolerance and control charts, acceptance and rejection testing, and an introduction to the design of experiments. RSK

Statistics, P. *Advances in Multivariate Statistical Analysis: Pillai Memorial Volume.* Ed: A.K. Gupta. Theory & Dec. Lib., Ser. B. D Reidel (US Distr: Kluwer Academic), 1987, xvi + 389 pp, $89. [ISBN: 90-277-2531-4] Twenty-one papers providing a cross-section of recent developments in multivariate statistical analysis. Dedicated to the memory of K.C. Sreedharan Pillai (1920-1985), it contains a short biography and a bibliography of his works. Note price. RSK

Statistics, T(17-18: 1), S, P*. *The Asymptotic Theory of Extreme Order Statistics, Second Edition.* Janos Galambos. Robert E Krieger, 1987, xv + 414 pp, $49.50. [ISBN: 0-89874-957-3] Revision of the author's 1978 *First Edition* published by Wiley (TR, December 1978). Rigorous presentation covering all known asymptotic models. Extensive bibliography, supplemented by a survey of the literature in each chapter. RSK

Statistics, T(17), P, L. *Kendall's Advanced Theory of Statistics, Fifth Edition of Volume 1: Distribution Theory.* Alan Stuart, J. Keith Ord. Oxford U Pr, 1987, xvi + 604 pp, $75. [ISBN: 0-19-520561-8] Major revision of the 1977 *Fourth Edition* of the first volume of this classic three-volume treatise, originally written by Kendall (1907-1983). The topics covered remain basically the same, but much new material has been added and terminology and proofs have been updated. RSK

Statistics, T(14-16: 1, 2), S, L. *Applied Statistics: Analysis of Variance and Regression, Second Edition.* Olive Jean Dunn, Virginia A. Clark. Prob. & Math. Stat. Wiley, 1987, xii + 445 pp, $36.95. [ISBN: 0-471-81269-2] Revision of the authors' 1974 text (TR, February 1975). Each chapter now includes a brief description of relevant BMDP and SAS computer programs. Also includes new material on repeated measure designs and the use of dummy variables in multiple regression and covariance analysis, and expanded material on variable selection and testing assumptions in multiple regression. RSK

Statistics, T(18: 1), P. *Asymptotic Distribution Theory in Nonparametric Statistics.* Manfred Denker. Adv. Lect. in Math. Friedr Vieweg & Sohn, 1985, vii + 204 pp, $18 (P). [ISBN: 3-528-08905-9] Treats three basic types of statistics: Hoeffding's U-statistics, differentiable statistical functionals, and statistics based on ranks. Concludes with a chapter on contiguity and efficiency. No exercises. RSK

Statistics, T(17: 1, 2), P. *Plane Answers to Complex Questions: The Theory of Linear Models.* Ronald Christensen. Texts in Stat. Springer-Verlag, 1987, xiv + 380 pp, $42. [ISBN: 0-387-96487-8] Designed "to rigorously illustrate the practical application of the projective approach to linear models." First half covers standard topics in regression analysis, analysis of variance and covariance; last part introduces various special topics such as residual analysis, variance component estimation, and log-linear models. RSK

Statistics, T(18: 1), S, P*. *Measurement Error Models.* Wayne A. Fuller. Prob. & Math. Stat. Wiley, 1987, xxiii + 440 pp, $44.95. [ISBN: 0-471-86187-1] Presents both theory and applications of models in which the explanatory variables are measured with error. Includes a number of real examples illustrating the procedures. Good bibliography. RSK

Statistics, S(16-17), P. *Design, Data, and Analysis by Some Friends of Cuthbert Daniel.* Ed: Colin L. Mallows. Prob. & Math. Stat. Wiley, 1987, xix + 380 pp, $29.95. [ISBN: 0-471-83937-X] Seventeen papers, three expository, two concerned with design issues, and the other twelve dealing primarily with the analysis of data, written in recognition of Daniel's great influence in this latter area. Written for "students and aspiring statistical consultants." Includes a brief outline of Daniel's career and a list of his publications. RSK

Statistics, P. *Contributions to the Theory and Application of Statistics: A Volume in Honor of Herbert Solomon.* Ed: Alan E. Gelfand. Academic Pr, 1987, xxviii + 544 pp, $59.95. [ISBN: 0-12-279450-8] Twenty papers, contributed by friends and colleagues, grouped into four areas where Solomon has made significant contributions: operations research and applied probability; distribution theory and geometric probability; applications in the areas of law and justice, medicine and psychology; and inference methodology. Includes a brief biographical sketch and a list of Solomon's publications. RSK

Statistics, T(16-17: 1), S, L. *Introduction to Statistical Inference.* Jack Carl Kiefer. Ed: Gary Lorden. Texts in Stat. Springer-Verlag, 1987, viii + 334 pp, $48. [ISBN: 0-387-96420-7] Text based on lecture notes developed by Kiefer (1924-1981) for a first course in statistical inference. Presents a mod-

ern decision-theoretic approach to inference, emphasizing the need to justify the use of a procedure by some criterion of goodness. RSK

Statistics, P. *The Population-Sample Decomposition Method.* A.M. Wesselman. Intern. Stud. in Econ. & Econometrics, V. 19. Kluwer Academic, 1987, 242 pp, $47.50. [ISBN: 90-247-3603-X] An approach which does not require all the model assumptions of classical statistical methods. The study aims to illustrate how the population-sample decomposition method is applicable to a much wider class of statistics. LCL

Programming, T(13-14: 1). *Pascal-SC: A Computer Language for Scientific Computation.* Gerd Bohlender, *et al.* Perspect. in Comput., V. 17. Academic Pr, 1987, ix + 292 pp, $34. [ISBN: 0-12-111155-5] Pascal-SC (scientific computation—implemented for Z80, 8088, and 68000 processors) is an extension of standard Pascal which allows greater accuracy in solving numerical problems, and has dynamic arrays and better string handling; chapters (among others) include standard Pascal, real floating-point arithmetic, strings and text processing, dynamic arrays, and modules; a syntax diagram appendix is included, along with numerous complete examples; useful as either a reference or text for a numerical-based Pascal course. RSF

Languages, P. *Lecture Notes in Computer Science-274: Functional Programming Languages and Computer Architecture.* Ed: Gilles Kahn. Springer-Verlag, 1987, vi + 470 pp, $34.60 (P). [ISBN: 0-387-18317-5] 24 papers from the third conference on the title subject, held in Portland, Oregon, September 1987. This conference focused particularly on implementation techniques for functional programming languages and on computer architectures to support the efficient execution of functional programs. DFA

Languages, P. *REDUCE: Software for Algebraic Computation.* Gerhard Rayna. Symbolic Computation. Springer-Verlag, 1987, ix + 329 pp, $29.80 (P). [ISBN: 0-387-96598-X] An introduction to REDUCE 3—a system for symbolic algebraic computation available on systems ranging in scale from an IBM PC to a Cray X/MP. Of special interest is a section containing several case studies illustrating the use of this symbolic computation package to solve nontrivial problems. AO

Languages, P. *Lecture Notes in Computer Science-276: ECOOP '87: European Conference on Object-Oriented Programming.* Ed: J. Bézivin, *et al.* Springer-Verlag, 1987, vi + 273 pp, $23.10 (P). [ISBN: 0-387-18353-1] Contains 25 papers on various aspects of object-oriented programming. AO

Computer Systems, P. *Lecture Notes in Computer Science-275: System Development and Ada.* Ed: A.N. Habermann, U. Montanari. Springer-Verlag, 1987, 305 pp, $25.70 (P). [ISBN: 0-387-18341-8] An interesting collection of papers presented at the workshop on software factories and Ada held in Capri, Italy, May 1986. Covers three general topics. First, design of software development environments, describing commercially available environments to ease production of very large software systems. Language-specific and mixed-language environments are presented. Secondly, methods for software development, addressing the issue of ensuring that implementations accurately reflect specifications. Finally, Ada compiler validation and specifying and testing Ada events are covered. PS

Theory of Computation, P. *Lecture Notes in Computer Science-273: A Connotational Theory of Program Structure.* James S. Royer. Springer-Verlag, 1987, 186 pp, $18 (P). [ISBN: 0-387-18253-5] This monograph is an outgrowth of a Ph.D. dissertation (SUNY at Buffalo) continuing development of a language-independent theory of program structure begun by Riccardi and Case; the central theme is the subclass, called acceptable numberings, of effective numberings; also stressed is "building" one control structure from others. It is very mathematical, with symbolism, numerous definitions, lemmas, propositions, and theorems; includes a bibliography, notation index, and definition index. RSF

Theory of Computation, P. *Lecture Notes in Computer Science-270: Computation Theory and Logic.* Ed: Egon Börger. Springer-Verlag, 1987, ix + 442 pp, $34.60 (P). [ISBN: 0-387-18170-9] Memorial volume for Dieter Rödding, a German logician interested in interaction of logic and computer science. Topics include recursion theory, automata theory, complexity theory, applications of logic in systems theory. KS

Theory of Computation, P. *Logic of Programming and Calculi of Discrete Design.* Ed: Manfred Broy. NATO ASI Ser. F, V. 36. Springer-Verlag, 1987, 413 pp, $75. [ISBN: 0-387-18003-6] Proceedings of a NATO Advanced Study Institute held in Marktoberdorf, Federal Republic of Germany, July 29-August 10, 1986. Contains fourteen papers on methodologies for the specification, design, and verification of programs. AO

Theory of Computation, T*(16-18: 1, 2), S, P, L. *Foundations of Logic Programming, Second, Extended Edition.* J.W. Lloyd. Symbolic Computation. Springer-Verlag, 1987, xii + 212 pp, $44.50. [ISBN: 0-387-18199-7] Expanded material in the *Second Edition* leans toward databases. As well, the class of programs considered is enlarged. Preliminaries are in the first chapter while the second chapter contains the declarative and procedural semantics of definite programs. Chapter three deals with normal programs and chapter four is new and treats unrestricted programs. The last chapter provides a theoretical foundation for deductive database systems. Chapter problems. References. Index. (*First Edition*, TR, October 1985.) RJA

Artificial Intelligence, P. *Natural Language Parsing Systems.* Ed: Leonard Bolc. Symbolic Computation. Springer-Verlag, 1987, xviii + 367 pp, $49.50. [ISBN: 0-387-17537-7] This collection of nine essays (fourteen authors) about current problems of natural language parsing in an artificial intelligence context

presents research results in greater detail than journal articles allow; each paper has an abstract and reference list; there is a volume subject index. RSF

Artificial Intelligence, P. *The Knowledge Frontier: Essays in the Representation of Knowledge.* Ed: Nick Cercone, Gordon McCalla. Symbolic Computation. Springer-Verlag, 1987, xxxv + 512 pp, $42. [ISBN: 0-387-96557-2] This collection of 17 essays (26 authors) about knowledge representation is an outgrowth of the IEEE Computer Special Issue on Knowledge Representation in 1983; six of the papers are new, 7 are updated, and 4 are unchanged; the papers are organized in six sections—Overview, Logic, Foundations, Organization, Reasoning, and Applications, giving a comprehensive treatment to this aspect of artificial intelligence. The preface, index, and reference list are extensive. RSF

Computer Science, P. *Lecture Notes in Computer Science-263: Advances in Cryptology—CRYPTO '86.* Ed: A.M. Odlyzko. Springer-Verlag, 1987, xi + 487 pp, $34.60 (P). [ISBN: 0-387-18047-8] Proceedings of an annual conference held at the University of California, Santa Barbara, devoted to cryptologic research. Thirty-six papers classified by data encryption, public-key cryptography, zero-knowledge proofs, secret-sharing methods, hardware systems, software systems, software protection, probabilistic methods, and other topics. LCL

Computer Science, P. *Lecture Notes in Computer Science-265: Analogical and Inductive Inference.* Ed: K.P. Jantke. Springer-Verlag, 1987, vi + 227 pp, $20.60 (P). [ISBN: 0-387-18081-8] Written versions of presentations given at the International Workshop on Analogical and Inductive Inference held in Wendisch-Rietz, GDR, October 6-10, 1986. Discussion centered around two basic approaches to learning algorithms: inductive inference and analogical reasoning. RJA

Computer Science, P. *Lecture Notes in Computer Science-264: Logic Programming '86.* Ed: Eiiti Wada. Springer-Verlag, 1987, 179 pp, $18 (P). [ISBN: 0-387-18024-9] Proceedings of the Fifth Logic Programming Conference held in Tokyo, Japan, June 23-26, 1986. RJA

Applications, P. *Reliability Data Bases.* Ed: Aniello Amendola, Alfred Z. Keller. D Reidel (US Distr: Kluwer Academic), 1987, xii + 398 pp, $72. [ISBN: 90-277-2549-7] Probabilistic techniques are increasingly used to assess safety, new system designs, and reliability of products. These applications require the availability of good dependable and relevant data about failure rates, incidents, performance, and so forth. These proceedings (ISPRA Course, October 1985) provide a comprehensive review of the state-of-the-art relating to the collection, processing, and use of such data. LCL

Applications, P. *Filtering for Stochastic Processes with Applications to Guidance, Second Edition.* Richard S. Bucy, Peter D. Joseph. Chelsea, 1987, xviii + 217 pp, $19.95. [ISBN: 0-8284-0326-0] Filtering a random signal process from observations corrupted by noise. A reprint of a 1968 edition which gives a detailed derivation of the Kalman-Bucy filter (TR, May 1969). Includes appendices which correct the *a priori* bounds and gives an up-to-date list of relevant references. CEC

Applications (Biological Science), P. *Mathematical Aspects of Hodgkin-Huxley Neural Theory.* Jane Cronin. Stud. in Math. Biology, V. 7. Cambridge U Pr, 1987, xi + 261 pp, $49.50. [ISBN: 0-521-33482-9] An introduction to the study of mathematical models of electrically active cells. Provides an account of the literature on the subject and covers in detail the Hodgkin-Huxley model for nerve conduction. Accessible to mathematicians with little or no background in physiology. RH

Applications (Cognitive Science), S, P, L.** *Brains, Machines, and Mathematics, Second Edition.* Michael A. Arbib. Springer-Verlag, 1987, xvi + 202 pp, $27. [ISBN: 0-387-96539-4] Revision (principally an expanded introduction) of a path-breaking book written in 1962 (published by McGraw-Hill in 1964) outlining a theory of automata to explain the processing of information in brain-like machines. Beginning with the McCulloch and Pitts model of neural networks and ending in Gödel's theorem, Arbib joins biology to mathematics in a prescient challenge to contemporary cognitive science. LAS

Applications (Electrical Engineering), T(15-16: 1). *Integrated Circuit Design.* Alan F. Murray, H. Martin Reekie. Springer-Verlag, 1987, xiv + 152 pp, $28. [ISBN: 0-387-91303-3] Designed as a textbook for an introductory course for electrical engineering or physics students. The final chapter provides a design exercise that can be assigned as a year-long project. AO

Applications (Electrical Engineering), T(16-17: 2), S, P, L. *Methods in Electromagnetic Wave Propagation.* D.S. Jones. Oxford Eng. Sci. Ser., V. 8. Clarendon Pr, 1987, $29.95 each (P). *Volume 1: Theory and Guided Waves,* 433 pp, [ISBN: 0-19-856189-X]; *Volume 2: Radiating Waves,* 476 pp. [ISBN: 0-19-856190-3] The objective of these volumes is to develop a suitable framework of theory and numerical analysis with applications to various aspects of propagation of electromagnetic waves. Numerous exercises are included. MU

Applications (Engineering), P. *Reliability Modelling and Applications.* Ed: A.G. Colombo, A.Z. Keller. D Reidel (US Distr: Kluwer Academic), 1987, x + 391 pp, $74. [ISBN: 90-277-2566-7] Proceedings of the ISPRA Course (November 1985) includes twenty papers grouped into five sections: systems reliability, availability and maintainability of systems, structural reliability, reliability of computing systems, human reliability. LCL

Applications (Engineering), P. *Lecture Notes in Mathematics-1279: Saint-Venant's Problem.* Dorin Ieşan. Springer-Verlag, 1987, viii + 162 pp, $16.30 (P). [ISBN: 0-387-18361-2] The problem of A.J.C.B. de Saint-Venant is to "determine the equilibrium of a homogeneous and isotropic linearly elastic cylinder,

loaded by surface forces distributed over its plane ends." These lecture notes present recent research on Saint-Venant's problem for anisotropic elastic bodies, on the relaxed Saint-Venant problem for heterogeneous elastic cylinders, and on Saint-Venant problems within the linearized theory of Cosserat elastic bodies. One hundred sixty references from Saint-Venant's original papers to the present. JK

Applications (Engineering), P. *Contact Mechanics.* K.L. Johnson. Cambridge U Pr, 1987, xi + 452 pp, $34.50 (P). [ISBN: 0-521-34796-3] Covers stresses and deformations which arise when two solid bodies come in contact, especially when the contact is localized due to dissimilar profiles (non-conformal). This induces a local stress concentration which can be studied using the method of superposition of point force solutions. MR

Applications (Engineering), P. *Lecture Notes in Engineering-9: The Complex Variable Boundary Element Method.* Theodore V. Hromadka II. Springer-Verlag, 1984, xi + 243 pp, $22 (P). [ISBN: 0-387-13743-2] Develops methods for approximating solutions to the Dirichlet and Neumann boundary-value problems by applying the Cauchy integral formula to the boundary element method. Includes a quick review of basic complex variable theory. BH

Applications (Fluid Mechanics), P. *Amorphous Polymers and Non-Newtonian Fluids.* Ed: Constantine Dafermos, J.L. Ericksen, David Kinderlehrer. Inst. for Math. & Its Applic., V. 6. Springer-Verlag, 1987, xii + 195 pp, $22. [ISBN: 0-387-96556-4] Ten papers from a workshop held at the Institute for Mathematics and Its Applications (IMA) in Minnesota during the 1984-85 special year on continuum physics and partial differential equations. LCL

Applications (Physics), P, L. *Three Hundred Years of Gravitation.* Ed: S.W. Hawking, W. Israel. Cambridge U Pr, 1987, xiii + 684 pp, $69.50. [ISBN: 0-521-34312-7] Sixteen chapters, separately authored, surveying cosmology, gravitation, string theory, and other theories descended from Newton's *Principia,* published just 300 years ago. Much more than a simple collection of papers, this comprehensive commemorative volume is a superb, well-planned exposition of the modern theory of gravitation. LAS

Applications (Physics), T(16-17: 1, 2), S, L. *Gravitational Physics of Stellar and Galactic Systems.* William C. Saslaw. Cambridge U Pr, 1987, xvii + 491 pp, $34.50 (P). [ISBN: 0-521-34975-3] Although primarily a theoretical treatise based on classical Newtonian gravity, some observational data is introduced for motivation. The text is relatively self-contained and accessible to anyone familiar with advanced calculus. Numerous interesting exercises are included. MU

Applications (Physics), P. *Symmetries of Maxwell's Equations.* W.I. Fushchich, A.G. Nikitin. Transl: John R. Schulenberger. Math. & Its Applic. D Reidel (US Distr: Kluwer Academic), 1987, xiv + 214 pp, $74. [ISBN: 90-277-2320-6] The power of Maxwell's equations lies largely in their symmetries, both geometric (Lie-group) and non-geometric (Lie-algebra). The former is classical; the latter is still being developed, and is the main subject here. The moral: Symmetry makes the world go 'round. BC

Applications (Simulation), P. *DEMOS: A System for Discrete Event Modelling on Simula.* G.M. Birtwistle. Springer-Verlag, 1987, 215 pp, $18 (P). [ISBN: 0-387-91301-7] An introduction to discrete event simulation modelling using DEMOS—a system which compliments SIMULA by providing elements which help beginners write simulations more quickly. Describes the basic DEMOS approach to model building as well as DEMOS descriptions of synchronization problems arising in discrete event simulations. Written as a teaching text for DEMOS, not as a reference. Tutorial style presents new features and offers an illustrative example of each. DEMOS programs' advantage over regular SIMULA programs: more easily written and understood without a thorough knowledge of SIMULA. Includes exercises and solutions. PS

Applications (Social Science), S, P, L. *Multidimensional Similarity Structure Analysis.* I. Borg, J. Lingoes. Springer-Verlag, 1987, xiv + 390 pp, $39. [ISBN: 0-387-96525-4] A class of models that represent similarity coefficients among a set of objects (e.g., correlation matrix) as distances in multidimensional space (two points are closer together when they are more correlated). The resulting picture is easier to assimilate than the table of coefficients. This book of case studies deals with all aspects of this subject, starting from scratch assuming no more than a high school background in mathematics. LCL

Reviewers

RJA: Richard J. Allen, St. Olaf; DFA: David F. Appleyard, Carleton; RB: Richard Brown, Carleton; LB-E: Lisa Brown-Ewing, Macalester; JNC: Judith N. Cederberg, St. Olaf; LC: Laura Chihara, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; RSF: Robert S. Fisk, St. Olaf; SG: Steven Galovich, Carleton; GG: George Gilbert, St. Olaf; BH: Bruce Hanson, St. Olaf; RH: Rhonda Hatcher, St. Olaf; RK: Roger Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; AM: Alan Magnuson, St. Olaf; SM: Steve McKelvey, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; MR: Matthew Richey, St. Olaf; AWR: A. Wayne Roberts, Macalester; GMS: G. Michael Schneider, Macalester; PS: Paul Schoening, St. Olaf; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MS: Myriam Steinback, Macalester; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; LW: Lynne Walling, St. Olaf; FLW: Frank L. Wolf, Carleton; MZ: Michael Zielinski, St. Olaf.

JOIN MAA TODAY!

Send this application to: MAA Membership Department
1529 18th Street, N.W.
Washington, D.C. 20036

Name ______________________________

Mailing Address ______________________________

______________________________ Zip __________

Employer/School ______________________________

Position (Rank) ______________________________

Employer's City/State ______________________________

Highest Earned Degree ______________ Year Degree Earned ______________

Institution Awarding Degree ______________________________

Month/Year of Birth ______________________________

Have you been a member of the MAA before? ☐ **Yes** ☐ **No**

Your membership may begin on July 1 or January 1. You may elect to receive any of the listed combinations of the three MAA journals. All members receive *Focus,* the MAA newsletter.

Please **circle** the appropriate box corresponding to your initial membership period and the selection of journals you wish to receive. (Rates are guaranteed for the indicated periods only.) Subscription prices are included with dues (see **).

THE AMERICAN MATHEMATICAL MONTHLY **(M)**
MATHEMATICS MAGAZINE **(G)**
THE COLLEGE MATHEMATICS JOURNAL **(J)**

Student Membership	**M**	**G**	**J**	**M+G**	**M+J**	**G+J**	**M+G+J**
1 year (Jan.-Dec. 1989)	$29.00	$23.00	$25.00	$36.00	$38.00	$31.00	$ 44.00
1½ year (July 1988-Dec. 1989)	$43.50	$34.50	$37.50	$54.00	$57.00	$46.50	$ 66.00
Regular Membership							
1 year (Jan.-Dec. 1989)	$48.00	$35.00	$39.00	$61.00	$65.00	$52.00	$ 78.00
1½ year (July 1988-Dec. 1989)	$72.00	$52.50	$58.50	$91.50	$97.50	$78.00	$117.00

*Student membership available to high school and undergraduate students and to students regularly enrolled in graduate study at least half time. Enclose letter from school or department official confirming status. Student rates apply to unemployed persons who are seeking employment.

**Annual dues include subscription prices as follows. Regular Member $26 (M), $17 (J), $13 (G), $3 *(Focus).* Student Member $13 (M), $9 (J), $7 (G), $1 *(Focus).*

Payment Enclosed $ __________ *U.S. Funds Only* (see blow)

METHOD OF PAYMENT ☐ Check Enclosed ☐ VISA ☐ MasterCard ☐ Bill me

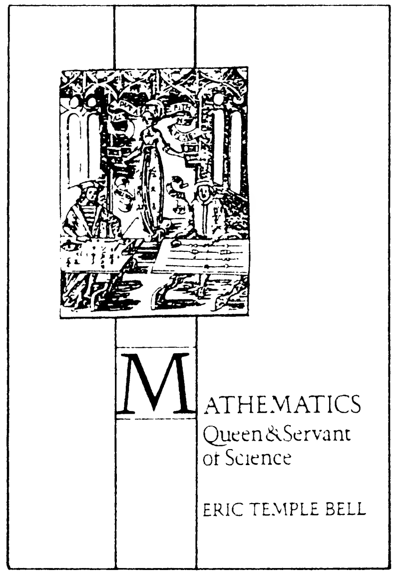
Card Number ______________________________ Expiration Date |__|__|__|__|
(month/year)

Interbank Number |__|__|__|__| *(MasterCard only—located above name on card)*

SIGNATURE ______________________________

MATHEMATICS: Queen and Servant of Science

by E. T. Bell



"Bell is a lively, stimulating writer, inoffensively crotchety and opinionated, with a good sense of historical circumstance, a fine impatience with humbug, a sound grasp of the entire mathematical scene, and a gift for clear and orderly explanation."

James R. Newman, *The World of Mathematics*

"Bell's writing style makes you feel as if you are a fortunate student spending time with a charming and gifted teacher. Not a history of mathematics, but a wide ranging exposition of ideas in the subject, the book is a classic and should be required reading for every teacher of mathematics."

Harold Jacobs, Grant High School, Van Nuys, California

"This book deserves a place in today's market. It is a much more popular work than most histories of the subject, and that is exactly what makes it accessible to students as well as to non-mathematicians. It is rewarding reading for . . . teachers and students of mathematics at all levels."

Morris Kline, Courant Institute

xv + 437 pp. Paperbound
Republished 1987.
ISBN-0-88385-446-3
List: $18.50
MAA Member: $14.00

This is an absorbing account of the development in pure and applied mathematics from the geometry of Euclid more than two-thousand years ago to some of the more recent developments in mathematical physics. Published 30 years ago, MATHEMATICS: QUEEN AND SERVANT OF SCIENCE remains a valuable tool for teachers, students, and all those enthusiastic about mathematics. The book includes 20 chapters on selected areas in pure and applied mathematics, including discussions on topics such as geometry, algebra, logic, probability, infinity, transformations, rings, groups, topology, and continuity. Also covered are subjects such as: Mersenne primes, quantum mechanics, and three-valued logic. The book is not a history of mathematics, but rather an enticing sampling of selected topics in mathematics and science.

The book contains fascinating information on the personalities and philosophies of the great mathematicians, plus perceptive resumes of their specific contributions to the moving force of this great science.

Order from:
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 95, Number 5 May 1988

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Dept. of Mathematics, Stanford University, Stanford, CA 94305 (address effective until May, 1988). Shorter articles about the teaching of mathematics should be sent to STAN WAGON and JOAN HUTCHINSON, Department of Mathematics, Smith College, Northampton, MA 01063. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to ANITA E. SOLOW, Dept. of Mathematics, Grinnell College, Grinnell, IA 50112; in geometry or topology to DENNIS DETURCK, Dept. of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to RICHARD LIBERA, Mathematical Sciences Department, University of Delaware, Newark, DE 19716. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Proposed problems (three copies) and solutions (two copies), both elementary and advanced to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, University of Calgary, Alberta, Canada T2N 1N4.

Reprint Permission: A. B. WILLCOX, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: MS. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Back Issues: P. and H. BLISS Co., Middletown, CT 06457.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $26 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $29. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $90 per year.

Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Assocation of America, 1529 Eighteenth Street, N.W., Washington, DC 20077-9564.

PRINTED IN THE UNITED STATES OF AMERICA

The "Prime Number Theorem" for the Periodic Orbits of a Bernoulli Flow

S. P. LALLEY
Department of Statistics, Purdue University, W. Lafayette, IN 47907

1. Introduction. In 1969 G. Margulis [**8**] published the statement of a remarkable result concerning the distribution of closed geodesics on a compact Riemann surface of curvature -1: if $N(x)$ is the number of closed geodesics with lengths not exceeding x, then

$$N(x) \sim e^x/x$$

as $x \to \infty$. Margulis' proof, unfortunately, has never been published in English. In 1976, Hejhal [**6**] gave a proof of Margulis' theorem based on the Selberg trace formula.

Since Margulis' announcement statistical regularities in the distributions of periodic orbits have been discovered for various flows. For a weakly mixing Axiom A flow restricted to a basic set, Parry and Pollicott [**10**] (following earlier work by Bowen [**1**], [**2**]) proved that the number of periodic orbits with minimal period not exceeding x is asymptotic to e^{hx}/hx as $x \to \infty$, where h is the topological entropy of the flow. Sarnak [**11**] proved a similar result for the horocycle flow.

These results closely resemble the prime number theorem. This is no accident: the proofs in [**6**], [**10**], and [**11**] all use suitable zeta functions together with some of the standard machinery of analytic number theory. In a recent paper Parry [**9**] has taken this approach one step further and produced an analogue of the Dirichlet density theorem for periodic orbits of Axiom A flows.

In this article I shall present a similar result for a very simple flow, the so-called Bernoulli flow. I shall use only elementary techniques of asymptotic analysis: no zeta functions, no Tauberian theorems, no heavy machinery. This approach has the advantage that it leads to some interesting results concerning the distributions of *individual* periodic orbits, results which have no analogues in analytic number theory. In another paper [**7**] I have shown that this elementary method can be adapted to the general Axiom A flow, giving similar results.

No knowledge of geometry or dynamical systems is necessary to understand this paper.

2. Symbolic flows and Bernoulli flows. Consider the map σ: $[0,1] \to [0,1]$ given by $\sigma x = \langle\langle 2x \rangle\rangle$, where $\langle\langle y \rangle\rangle$ denotes the fractional part of y (i.e., $\langle\langle y \rangle\rangle = y - [[y]]$, where $[[y]]$ is the greatest integer in y). The map σ is called the *shift* because of what it does to the binary expansion of x: if $x = .x_1x_2x_3\ldots$, then $\sigma x = .x_2x_3x_4\ldots$.

Let f: $[0,1] \to (0,\infty)$ be a continuously differentiable function. (More generally, one may take f to be piecewise C^1 with discontinuities at dyadic rationals.) The *symbolic flow under f* (sometimes called the *f-suspension* or the *special flow under f*)

After a brief career wrestling alligators in carnivals, Lalley took up the study of probability and statistics. He wrote his Ph.D. dissertation on sequential testing at Stanford University under the tutelage of David Siegmund. He was a member of the statistics faculty at Columbia University from 1980 until 1986, and is now at Purdue University.

takes place on the region $\Sigma_f = \{(x, s): 0 \leq x \leq 1 \text{ and } 0 \leq s < f(x)\}$ under the graph of f. The dynamics are as follows: starting at any point $(x, s) \in \Sigma_f$, a particle moves upward along the vertical line segment from (x, s) to $(x, f(x))$ at unit velocity; upon reaching the "ceiling" at $(x, f(x))$ the particle jumps instantaneously to the "floor" at $(\sigma x, 0)$, then proceeds at unit speed up the vertical line segment from $(\sigma x, 0)$ to $(\sigma x, f(\sigma x))$, jumping again upon reaching the ceiling, etc. If $T_t(x, s)$ denotes position at time t of a particle started at (x, s) initially, then

$$T_t(x, s) = \begin{cases} (x, s + t) & \text{if } s + t < f(x) \\ (\sigma x, 0) & \text{if } s + t = f(x) \end{cases}$$

and

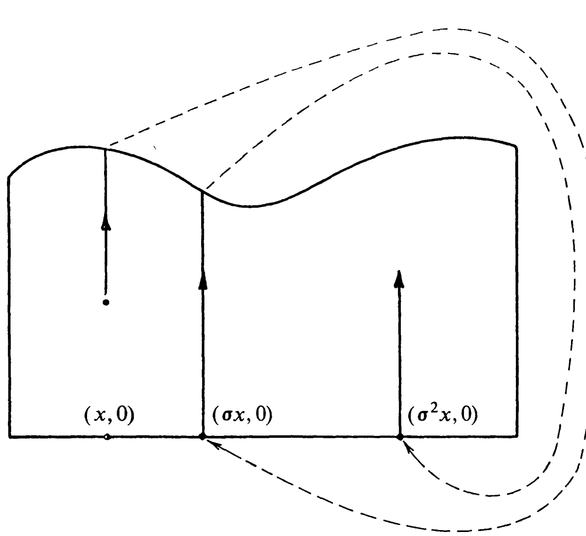
$$T_t \circ T_s = T_{t+s}.$$

A *Bernoulli flow* is a symbolic flow under a function f of the form

$$f(x) = 1\left\{x \leq \frac{1}{2}\right\} + \beta 1\left\{x > \frac{1}{2}\right\}.$$

For many interesting flows, including geodesic flows on compact manifolds of negative curvature and, more generally, Axiom A flows, questions about the ergodic behavior of trajectories may be reduced to similar questions about trajectories of symbolic flows. This is because the trajectories of such flows may be "coded" into binary expansions in a continuous and (more or less) one-to-one fashion. The idea is that if the phase space is partitioned into finitely many "boxes" then trajectories of the flow are more or less uniquely determined by the sequences of boxes the trajectories enter and leave. The matter of choosing a suitable partition is quite difficult: cf. Bowen **[2]** and Series **[12]**.

The key to the study of periodic orbits of the symbolic flow under f is that the periodic orbits are *precisely* those orbits which intersect the "floor" (i.e., x-axis) at points $(x, 0)$ for which x has a periodic binary expansion. Observe that if x has a binary expansion with (minimal) period m then the orbit through $(x, 0)$ intersects the floor of Σ_f at $(x, 0), (\sigma x, 0), (\sigma^2 x, 0), \ldots, (\sigma^{m-1} x, 0)$, and at no other points.



Furthermore, the (minimal) period of such an orbit is

$$f(x) + f(\sigma x) + \cdots + f(\sigma^{m-1}x). \tag{2.1}$$

Observe that for the Bernoulli flow the quantity (2.1) is just

$$m + (\beta - 1)\sum_{i=1}^{m} x_i,$$

where $.x_1x_2x_3\ldots$ is the binary expansion of x.

The main result of this article is an asymptotic formula for the number of periodic orbits of a Bernoulli flow with $\beta > 1$ and β irrational (cf. (9.4) below). This will be derived in sections 4–9. First, however, I shall look at a special case.

3. A special case. For the symbolic flow under the constant function $f \equiv 1$ the counting problem is almost trivial. The number of periodic binary expansions with minimal period dividing n is 2^n, so by the Möebius inversion formula the number of n-periodic binary expansions is

$$\sum_{d|n} 2^d \mu\left(\frac{n}{d}\right),$$

μ being the Möebius function. Each periodic orbit of the flow with period n corresponds to a set of n such n-periodic binary expansions: therefore, the number of periodic orbits of the flow with period n is

$$n^{-1}\sum_{d|n} 2^d \mu\left(\frac{n}{d}\right) \sim 2^n/n \tag{3.1}$$

as $n \to \infty$. This agrees in form with the results of Margulis and Parry and Pollicott.

Now consider the ensemble of periodic binary expansions with minimal period dividing n. By the weak law of large numbers for Bernoulli trials, for "most" periodic binary expansions with period dividing n the proportion of ones is near $1/2$, when n is large. Hence "most" periodic orbits of the flow with period dividing n spend approximately half of the time on the right half of Σ_f and the rest on the left half of Σ_f.

Much more can be said. Let $e_1e_2 \cdots e_m$ be a finite sequence of zeros and ones. Then the weak law of large numbers implies that for "most" periodic binary expansions with period dividing n, the proportion of times the pattern $e_1e_2 \cdots e_m$ appears in the expansion is near 2^{-m}, provided n is large. Thus most periodic orbits of the flow with period dividing n spend about 2^{-m} of the time in each rectangle with vertices $(k2^{-m},0)$, $(k2^{-m},1)$, $((k+1)2^{-m},1)$, and $((k+1)2^{-m},0)$, $k = 0,1,\ldots,2^m - 1$. Since the number of periodic orbits with period n is asymptotic to the number of periodic orbits with period dividing n (cf. (3.1)), it follows that *most periodic orbits with period n are nearly uniformly distributed on the square* Σ_f.

It is tempting to conjecture that this result is true for all symbolic flows. The truth, however, is more subtle, as will be seen.

4. Counting orbits of the Bernoulli flow. Consider now the Bernoulli flow described in section 2, i.e., the symbolic flow under

$$f(x) = 1\left\{x \le \frac{1}{2}\right\} + \beta 1\left\{x > \frac{1}{2}\right\}.$$

For definiteness, I assume $\beta > 1$. In addition, I assume that β is *irrational*: the reason for this will be seen later (cf. section 7). Notice, however, that if β were not irrational, then $\beta = k/l$ for certain relatively prime integers k and l and the lengths of all periodic orbits would be integer multiples of $1/l$.

Recall that a periodic orbit of the Bernoulli flow is one which passes through the floor of Σ_f at a point $(x, 0)$ such that x has a periodic binary expansion, and that the period of such an orbit is $\sum_{i=1}^{m} f(\sigma^i x)$, m being the period of the binary expansion. Let $\varepsilon > 0$ be sufficiently small that

$$\varepsilon_* = \varepsilon/(\beta - 1) < 1. \tag{4.1}$$

For $a > 0$, define

$N(a; \varepsilon) = \#\{$*periodic orbits with period between* a *and* $a + \varepsilon\}$,

$N_m^*(a; \varepsilon) = \#\{$*periodic binary expansions* $.x_1 x_2 x_3 \cdots$ *with period* m *such that*

$$a - m < (\beta - 1) \sum_{i=1}^{m} x_i \le a - m + \varepsilon\Big\},$$

$N_m(a; \varepsilon) = \#\{$*periodic binary expansions* $.x_1 x_2 x_3 \cdots$ *with period dividing* m *such that*

$$a - m \le (\beta - 1) \sum_{i=1}^{m} x_i < a - m + \varepsilon\Big\}.$$

Then

$$N(a; \varepsilon) = \sum_{m=1}^{\infty} N_m^*(a; \varepsilon)/m$$

and

$$N_m(a; \varepsilon) = \sum_{d|m} N_d^*(ad/m; \varepsilon d/m).$$

The Möebius inversion formula implies that

$$N_m^*(a; \varepsilon) = \sum_{d|m} \mu\left(\frac{m}{d}\right) N_d(ad/m; \varepsilon d/m);$$

consequently,

$$N(a; \varepsilon) = \sum_{m=1}^{\infty} m^{-1} \sum_{d|m} \mu\left(\frac{m}{d}\right) N_d(ad/m; \varepsilon d/m). \tag{4.2}$$

The reason for representing $N(a; \varepsilon)$ this way is that the quantities $N_m(a; \varepsilon)$ may be easily calculated in terms of the binomial coefficients. Suppose there is an integer $k \ge 0$ such that $a - m < (\beta - 1)k \le a - m + \varepsilon$; notice that there can only be one such integer, because of (4.1). Then $N_m(a; \varepsilon)$ is the number of distinct sequences $x_1 x_2 \cdots x_m$ of zeros and ones with exactly k ones, so

$$N_m(a; \varepsilon) = \binom{m}{k}. \tag{4.3}$$

If there is no integer k satisfying $a - m < (\beta - 1)k \le a - m + \varepsilon$, then

$$N_m(a; \varepsilon) = 0. \tag{4.4}$$

It is worth noting that the numbers $N_m(a;\, \varepsilon)$ have a "probabilistic" interpretation. If S_m is a random variable with the binomial $(m, 1/2)$ distribution (i.e., S_m is the number of heads in m independent tosses of a fair coin), then

$$N_m(a;\, \varepsilon) = 2^m Pr\{a - m < (\beta - 1)S_m \le a - m + \varepsilon\},$$

thus the "asymptotics" of the numbers $N_m(a;\, \varepsilon)$ are tied up with some of the classical limit laws of probability theory.

Equations (4.2)–(4.4) provide an explicit representation for $N(a;\, \varepsilon)$ in terms of the binomial coefficients and the Möbius function. Unfortunately, this representation is not very illuminating, nor is it very useful for exact computations. In the next four sections I shall use methods of asymptotic analysis to obtain an approximate formula for $N(a;\, \varepsilon)$.

5. Asymptotic analysis: preliminaries. The series (4.2) contains two sums. I shall argue that the sum $\Sigma_{d|m}$ is dominated by the term $d = m$.

Suppose $d|m$. Let $.x_1^{(i)}x_2^{(i)}x_3^{(i)}\ldots$, $i = 1, 2, \ldots, m/d$, be periodic binary expansions each with period dividing d. One may form a periodic binary expansion $.x_1x_2x_3\ldots$ with period dividing m by splicing together d-blocks of the binary expansions $.x_1^{(i)}x_2^{(i)}x_3^{(i)}\ldots$ as follows:

$$.x_1x_2x_3\ldots = .x_1^{(1)}x_2^{(1)}\ldots x_d^{(1)}x_1^{(2)}x_2^{(2)}\ldots x_d^{(2)}\ldots x_d^{(m/d)}x_1^{(1)}\ldots.$$

This construction is one-to-one: each distinct set of binary expansions $.x_1^{(i)}x_2^{(i)}\ldots$, $i = 1, 2, \ldots, m/d$, gives rise to a distinct expansion $.x_1x_2x_3\ldots$. Furthermore, if each sequence $x_1^{(i)}x_2^{(i)}\ldots x_d^{(i)}$ has the property that

$$ad/m - d < (\beta - 1)\sum_{j=1}^{d} x_j^{(i)} \le ad/m - d + \varepsilon d/m,$$

then

$$a - m < (\beta - 1)\sum_{j=1}^{m} x_j \le a - m + \varepsilon.$$

It therefore follows that

$$N_d(ad/m;\, \varepsilon d/m)^{m/d} \le N_m(a;\, \varepsilon). \tag{5.1}$$

If $d|m$ and $d < m$ then $m/d \ge 2$. Since m has at most m nonnegative divisors it follows from (5.1) that

$$\left|\sum_{\substack{d|m\\ d\ne m}} \mu\left(\frac{m}{d}\right)N_d(ad/m;\, \varepsilon d/m)\right| \le mN_m(a;\, \varepsilon)^{1/2} \tag{5.2}$$

In subsequent sections I shall show that the sum $\Sigma_m N_m(a;\, \varepsilon)/m$ and the largest terms in this sum both grow at an *exponential* rate as $a \to \infty$. Hence the terms represented in (5.2) grow at a smaller exponential rate. In any case (5.1) and (5.2) imply

$$N(a;\, \varepsilon) = \sum_{m=1}^{\infty} \{m^{-1}N_m(a;\, \varepsilon) + C_m(a;\, \varepsilon)\}, \tag{5.3}$$

where

$$|C_m(a;\,\varepsilon)| \le N_m(a;\,\varepsilon)^{1/2}. \tag{5.4}$$

Recall that $N_m(a;\,\varepsilon)$ is the number of periodic binary expansions $.x_1x_2x_3\ldots$ with period dividing m such that

$$(a-m)/(\beta-1) < \sum_{i=1}^{m} x_i \le (a-m)/(\beta-1) + \varepsilon_*$$

($\varepsilon_* = \varepsilon/(\beta-1) < 1$ by (4.1)). The sum $\Sigma_{i=1}^m x_i$ is a nonnegative integer no larger than m. Since the difference between the extreme sides of the double inequality is $\varepsilon_* < 1$, there is at most one possible value for $\Sigma_{i=1}^m x_i$; furthermore, there is an integer between $(a-m)/(\beta-1)$ and $(a-m)/(\beta-1)+\varepsilon_*$ iff the fractional part of $(a-m)/(\beta-1)$ is at least $1-\varepsilon_*$. In order that this integer lie between zero and m it is necessary that

$$a/\beta \le m \le a+\beta-1.$$

It now follows from (5.3) and (5.4) that

$$\begin{aligned} N(a;\,\varepsilon) = \sum_{a/\beta\le m\le a+\beta-1} &\left(\binom{m}{k_m}\Big/ m + C_m(a;\,\varepsilon)\right) \\ &1\{\langle\langle (a-m)/(\beta-1)\rangle\rangle \ge 1-\varepsilon_*\}, \end{aligned} \tag{5.5}$$

where

$$k_m = \left[\left[(a-m)(\beta-1)^{-1}\right]\right] + 1 \tag{5.6}$$

and

$$|C_m(a;\,\varepsilon)| \le \binom{m}{k_m}^{1/2}. \tag{5.7}$$

6. Asymptotic analysis: applying Stirling's formula. The binomial coefficients in the series (5.5) may be handled by Stirling's formula and a related inequality (cf., e.g., Feller [**3**, section II.9]):

$$n! \sim n^n e^{-n}(2\pi n)^{1/2} \text{ as } n \to \infty,$$

$$n^n e^{-n} \le n! \le (n+1)^{n+1}e^{-n} \qquad \forall n = 1,2,\ldots.$$

The double inequality and the elementary relation $(1+m^{-1})^m \le e$ imply

$$\binom{n}{k} \le (n+1)e^{nH(k/n)+1} \qquad \forall n, k \ge 0,$$

where H is the Shannon entropy function

$$\begin{aligned} H(p) &= -p\log p - (1-p)\log(1-p), \qquad 0<p<1, \\ H(0) &= H(1) = 0; \end{aligned}$$

Stirling's formula implies that if $n\to\infty$, $k\to\infty$, and $n-k\to\infty$, then

$$\binom{n}{k} \sim e^{nH(k/n)}((2\pi n(k/n)(1-k/n)))^{-1/2}.$$

Consider now the quantities $\binom{m}{k_m}$ occurring in the sum (5.5). By the preceding remarks,

$$\binom{m}{k_m} \le (m+1)\exp\{mH(k_m/m)+1\} \quad \forall m, \tag{6.1}$$

and

$$\binom{m}{k_m} \sim \exp\{mH(k_m/m)\}\left(2\pi m^{-1}k_m(m-k_m)\right)^{-1/2} \tag{6.2}$$

provided both k_m and $m - k_m$ are large.

Define

$$J(p) = pH\left((1-p)p^{-1}(\beta-1)^{-1}\right), \qquad \beta^{-1} \le p \le 1,$$
$$J(p) = 0, \qquad p \notin [\beta^{-1}, 1],$$
$$Q(p) = H'\left((1-p)p^{-1}(\beta-1)^{-1}\right), \qquad \beta^{-1} \le p \le 1.$$

It follows from (5.6) that

$$k_m/m = (am^{-1}-1)(\beta-1)^{-1} + m^{-1}\left\{1 - \left\langle\left\langle (a-m)(\beta-1)^{-1}\right\rangle\right\rangle\right\}.$$

Consequently, since $H(p)$ is continuous for $p \in [0,1]$,

$$mH(k_m/m) \sim aJ(m/a) \tag{6.3}$$

as $a \to \infty$, uniformly for $a\beta^{-1} \le m \le a + \beta - 1$; moreover, since H is C^∞ on $(0,1)$, if m/a is restricted to a compact subset of $(\beta^{-1}, 1)$, then

$$\begin{aligned} mH(k_m/m) &= aJ(m/a) \\ &\quad + Q(m/a)\left\{1 - \left\langle\left\langle (am^{-1}-1)(\beta-1)^{-1}\right\rangle\right\rangle\right\} \\ &\quad + o(1), \end{aligned} \tag{6.4}$$

where $o(1)$ converges uniformly to zero as $a \to \infty$. It now follows from (6.1) and (6.3) that

$$\binom{m}{k_m} \le me^{aJ(m/a)}e^{o(a)} \quad \forall a\beta^{-1} \le m \le a + \beta - 1 \tag{6.5}$$

where $o(a)$ is independent of m and $\lim_{a\to\infty} o(a)/a = 0$; and it follows from (6.2) and (6.4) that

$$\begin{aligned} \binom{m}{k_m} &\sim e^{aJ(m/a)}\exp\left\{Q(m/a)\left\{1 - \left\langle\left\langle (a-m)(\beta-1)^{-1}\right\rangle\right\rangle\right\}\right\} \\ &\quad \times \left(2\pi m r(m/a)(1 - r(m/a))\right)^{-1/2}, \end{aligned} \tag{6.6}$$

where

$$r(p) = (1-p)p^{-1}(\beta-1)^{-1}, \qquad p \in (\beta^{-1}, 1), \tag{6.7}$$

as $a \to \infty$, uniformly for m/a in any compact subset of $(\beta^{-1}, 1)$.

It is apparent from (6.5) and (6.6) that the exponential growth rates of the binominal coefficients in the sum (5.5) are governed by the function $J(p)$, $p \in [\beta^{-1}, 1]$. Now $J(\beta^{-1}) = J(1) = 0$, since $H(0) = H(1)$, and $J(p) > 0$ for $p \in$

$(\beta^{-1}, 1)$, since $H(p) > 0$ for $p \in (0,1)$. Since J is continuous, it follows that J attains a maximum value in $(\beta^{-1}, 1)$. Routine calculations give

$$J''(p) = H''(r(p))/\left(p^3(\beta-1)^2\right), \tag{6.8}$$

$$H''(r) = -\left(r(1-r)\right)^{-1}, \tag{6.9}$$

hence J is strictly concave on $[\beta^{-1}, 1]$, and achieves its maximum value uniquely at some $p_\beta \in (\beta^{-1}, 1)$. Thus

$$J(p_\beta) = \max_{\beta^{-1} \le p \le 1} J(p) > 0, \tag{6.10}$$

$$J''(p_\beta) < 0. \tag{6.11}$$

Together with (5.5), (6.5), and (6.6), this suggests that $J(p_\beta)$ is the exponential growth rate of $N(a; \varepsilon)$. I shall prove this, and more, in section 8.

Note. It may be shown that the exponential growth rate $J(p_\beta)$ is the "topological entropy" of the Bernoulli flow, the topological entropy being the maximum measure-theoretic entropy achieved by an invariant measure. The invariant probability measure μ_β which maximizes entropy may be described as follows. Let $X = .X_1X_2X_3\ldots$, where $X_1, X_2, \ldots$ are i.i.d. Bernoulli $(r(p_\beta))$, and let Y be uniformly distributed on $[0, f_\beta(X)]$, where $f_\beta(x) = 1\{x \le 1/2\} + \beta 1\{x > 1/2\}$; then μ_β is the distribution of the point (X, Y). (For general results subsuming these, cf. Lalley [7].)

7. Asymptotic analysis: rapidly oscillating terms. Asymptotic analysis of the series (5.5) is complicated by the presence of factors depending on the quantity $\langle\langle (a-m)(\beta-1)^{-1}\rangle\rangle$, which cause rapid oscillations in the terms of the series. There is, however, a high degree of regularity in these oscillations, due to the fact that $1/(\beta-1)$ is irrational.

Call an increasing sequence $\{n_j\}$ of integers *k-syndetic* if $n_{j+1} - n_j \le k$ for all j (the terminology is adapted from Furstenberg [**5**, section 1.4]). For real numbers γ, $\delta \ne 0$, $0 < \varepsilon_* < 1$, let $\{n_j(\gamma, \delta, \varepsilon_*)\}$ be the increasing sequence of nonnegative integers n satisfying

$$\langle\langle \gamma + n\delta \rangle\rangle \ge 1 - \varepsilon_*.$$

LEMMA 1. *For any $0 < \varepsilon_* < 1$ and any irrational δ there exists an integer $k = k(\delta, \varepsilon_*) > 0$ such that for any $\gamma \in \mathscr{R}$, the sequence $\{n_j(\gamma, \delta, \varepsilon_*)\}$ is k-syndetic.*

Proof. Since the unit interval $[0,1]$ is compact the sequence $\langle\langle n\delta\rangle\rangle$, $n = 1, 2, \ldots$, has a convergent subsequence. Hence, there exist integers $n' < n''$ such that $|\langle\langle n'\delta\rangle\rangle - \langle\langle n''\delta\rangle\rangle| < \varepsilon_*/2$. Let $n = n'' - n'$; then either $\langle\langle n\delta\rangle\rangle < \varepsilon_*/2$ or $\langle\langle n\delta\rangle\rangle > 1 - \varepsilon_*/2$, and $\langle\langle n\delta\rangle\rangle \ne 0$ or 1, because δ is irrational. In either case, there exists a finite integer m such that every $x \in [0,1]$ is within $\varepsilon_*/2$ of some point in the finite sequence $\langle\langle n\delta\rangle\rangle, \langle\langle 2n\delta\rangle\rangle, \ldots, \langle\langle mn\delta\rangle\rangle$. Consequently, there exist nonnegative integers k_1, k_2 such that $\langle\langle k_1\delta\rangle\rangle < \varepsilon_*/2$ and $\langle\langle k_2\delta\rangle\rangle > 1 - \varepsilon_*/2$. Let $k = \max(k_1, k_2)$; then for any $\gamma \in \mathscr{R}$ the sequence $\{n_j(\gamma, \delta, \varepsilon_*)\}$ is k-syndetic, because if $\langle\langle x\rangle\rangle > 1 - \varepsilon_*$, then $\langle\langle x + k_1\delta\rangle\rangle > 1 - \varepsilon_*$ or $\langle\langle x + k_2\delta\rangle\rangle > 1 - \varepsilon_*$. $\diamondsuit$

LEMMA 2. *Let f: $[0,1] \to \mathscr{R}$ be continuous. Then for any irrational $\delta \in \mathscr{R}$, any $0 < \varepsilon_* < 1$, and any $\gamma \in \mathscr{R}$,*

$$\lim_{m\to\infty} m^{-1} \sum_{n=1}^{m} f(\langle\langle \gamma + n\delta \rangle\rangle) 1\{\langle\langle \gamma + n\delta \rangle\rangle \ge 1 - \varepsilon_*\}$$
$$= \int_{1-\varepsilon_*}^{1} f(t)\, dt. \tag{7.1}$$

Furthermore, this convergence holds uniformly in γ for each fixed $f \in C[0,1]$, $0 < \varepsilon_* < 1$, and irrational δ.

Proof. Weyl's equidistribution theorem (cf. Feller [**4**, section VIII, 4] or Furstenberg [**5**, section 3.3]) states that for any continuous g: $[0,1] \to \mathscr{R}$ such that $g(0) = g(1)$,

$$\lim_{m\to\infty} m^{-1} \sum_{n=1}^{m} g(\langle\langle n\delta \rangle\rangle) = \int_0^1 g(t)\, dt.$$

For any $\gamma \in \mathscr{R}$ the function $g_\gamma(t) = g(\langle\langle \gamma + t \rangle\rangle)$ is a continuous function on $[0,1]$ satisfying $g_\gamma(0) = g_\gamma(1)$ and $\int_0^1 g_\gamma = \int_0^1 g$, hence by Weyl's theorem,

$$\lim_{m\to\infty} m^{-1} \sum_{n=1}^{m} g(\langle\langle \gamma + n\delta \rangle\rangle) = \int_0^1 g(t)\, dt. \tag{7.2}$$

It is easily verified, using the uniform continuity of g, that this convergence holds uniformly in γ.

The function $f(t)1\{t \ge 1 - \varepsilon_*\}$ may be approximated from above and from below by continuous functions $g(t)$ satisfying $g(0) = g(1)$. Thus (7.1) follows from (7.2) and the uniformity in γ of (7.1) follows from the uniformity in γ of (7.2). ◇

8. Asymptotic analysis: reduction to a Gaussian integral. Using the results of sections 6 and 7, I shall evaluate the sum (5.5) asymptotically as $a \to \infty$. There will be three main steps in the analysis. First, I will show that only those terms for which $m/a \approx p_\beta$ (cf. (6.10)) contribute appreciably to the sum. Second, I will use Lemma 2 to dispose of the rapidly oscillating factors in the terms of the sum. Finally, I will show that the modified sum over the region $m/a \approx p_\beta$, with the rapid oscillations removed, may be approximated by a Gaussian integral.

Step 1. Inequalities (5.7) and (6.5) give upper bounds on the terms of the sum (5.5): to wit,

$$\left(\binom{m}{k_m}\Big/ m + C_m(a;\varepsilon)\right) 1\left\{\langle\langle (a-m)(\beta-1)^{-1} \rangle\rangle \ge 1 - \varepsilon_*\right\} \le e^{aJ(m/a)+o(a)} \tag{8.1}$$

for all m satisfying $a\beta^{-1} \le m \le a + \beta - 1$. Recall that $J(p)$ achieves its maximum value *uniquely* at $p = p_\beta$; since $J(p)$ is continuous it follows that for every $\delta > 0$, $\max_{|p-p_\beta|\ge\delta} J(p) < J(p_\beta)$. I will argue that there is at least one nonzero term in the sum (5.5) whose logarithm is approximately $aJ(p_\beta)$; hence, only those terms in (5.5) such that $|m/a - p_\beta| < \delta$ contribute appreciably to the sum.

By Lemma 1 there is an integer k such that for all $a > 0$ the sequence of integers m satisfying $\langle\langle (a-m)(\beta-1)^{-1} \rangle\rangle \ge 1 - \varepsilon_*$ is k-syndetic. Hence, there exists a

nonzero term in the sum (5.5) such that $|m/a - p_\beta| \le k/a$. But (5.7) and (6.6) imply that this term satisfies

$$\left(\binom{m}{k_m}\Big/ m + C_m(a;\varepsilon)\right) \sim e^{aJ(p_\beta)+o(a)}. \tag{8.2}$$

Now there are at most $0(a)$ nonzero terms in the sum (5.5), and they all satisfy (8.1). Moreover, there is at least one nonzero term satisfying (8.2). Since $J(p)$ is bounded away from $J(p_\beta)$ for $|p - p_\beta| \ge \delta$, it follows that for every $\delta > 0$, as $a \to \infty$,

$$\begin{aligned} N(a;\varepsilon_*) \sim \sum_{m:\,|ma^{-1}-p_\beta|<\delta} &\left(\binom{m}{k_m}\Big/ m + C_m(a;\varepsilon)\right) \\ &\times 1\left\{\left\langle\left\langle (a-m)(\beta-1)^{-1}\right\rangle\right\rangle \ge 1-\varepsilon_*\right\}. \end{aligned} \tag{8.3}$$

Notice that (8.1) and (8.2) also yield

$$\lim_{a\to\infty} a^{-1}\log N(a;\varepsilon_*) = J(p_\beta).$$

Step 2. Next, consider those terms in the sum (8.3) for which $|ma^{-1} - p_\beta| \le za^{-1/2}$, where $z > 0$ is a large constant. By (5.7) and (6.6),

$$\begin{aligned} &\left(\binom{m}{k_m}\Big/ m + C_m(a;\varepsilon)\right) 1\left\{\left\langle\left\langle (a-m)(\beta-1)^{-1}\right\rangle\right\rangle \ge 1-\varepsilon_*\right\} \\ &\quad\sim e^{aJ(m/a)}\exp\left\{Q(m/a)\left\{1-\left\langle\left\langle (a-m)(\beta-1)^{-1}\right\rangle\right\rangle\right\}\right\} \\ &\quad\times\left(2\pi a^3 r(m/a)(1-r(m/a))\right)^{-1/2}(m/a)^{-3/2} \\ &\quad\times 1\left\{\left\langle\left\langle (a-m)(\beta-1)^{-1}\right\rangle\right\rangle \ge 1-\varepsilon_*\right\} \end{aligned} \tag{8.4}$$

uniformly for m in the range $|ma^{-1} - p_\beta| \le \delta$. The expression on the right has a rapidly oscillating factor and a slowly oscillating factor. I will argue that the rapidly oscillating factor may be replaced by an average.

Let m_1, m_2 be any two integers in the range $|ma^{-1} - p_\beta| \le z/a^{1/2}$ such that $|m_1 - m_2| \le \log a$. Since $J'(p_\beta) = 0$, Taylor's theorem implies

$$|aJ(m_1/a) - aJ(m_2/a)| \le C(\log a)/a^{1/2}$$

for a suitable constant C (which may depend on z but not on a). Also, since $r(p)$ is continuous, $r(m/a) \sim r(p_\beta)$ for $|ma^{-1} - p_\beta| \le za^{-1/2}$. Therefore, the factor

$$e^{aJ(m/a)}\left(2\pi a^3 r(m/a)(1-r(m/a))\right)^{-1/2}(m/a)^{-1}$$

is nearly constant as m ranges over any interval of length $\log a$ contained in $|ma^{-1} - p_\beta| \le za^{-1/2}$.

The function $Q(p)$ is continuous, hence $Q(m/a) \sim Q(p_\beta)$ uniformly for $|ma^{-1} - p_\beta| \le za^{-1/2}$. Since $e^{Q(p_\beta)t}$ is a continuous function of t, it follows from Lemma 2 that the sum of

$$\exp\left\{Q(m/a)\left\{1-\left\langle\left\langle (a-m)(\beta-1)^{-1}\right\rangle\right\rangle\right\}\right\}$$

$$1\left\{\left\langle\left\langle (a-m)(\beta-1)^{-1}\right\rangle\right\rangle \ge 1-\varepsilon_*\right\}$$

over any interval of length [[log a]] contained in $|ma^{-1} - p_\beta| \le za^{-1/2}$ is asymptotic to

$$(\log a)\int_{1-\varepsilon_*}^{1} e^{Q(p_\beta)(1-t)}\,dt$$

as $a \to \infty$. Furthermore, this holds uniformly over all intervals of length [[log a]] contained in $|ma^{-1} - p_\beta| \le za^{-1/2}$, by the uniformity in Lemma 2.

Since the sum over the range $|ma^{-1} - p_\beta| \le za^{-1/2}$ may be broken up into sums over intervals of length [[log a]], the arguments of the preceding paragraphs show that as $a \to \infty$,

$$\sum_{m:\,|ma^{-1}-p_\beta|\le za^{-1/2}} \left(\binom{m}{k_m}\Big/m + C_m(a;\varepsilon)\right)1\left\{\left\langle\left\langle (a-m)(\beta-1)^{-1}\right\rangle\right\rangle \ge 1-\varepsilon_*\right\}$$

$$\sim \sum_{m:\,|ma^{-1}-p_\beta|\le za^{-1/2}} \left\{e^{aJ(m/a)}\left(2\pi a^3 r(m/a)(1-r(m/a))\right)^{-1/2}(m/a)^{-3/2} \times \int_0^{\varepsilon_*} e^{Q(p_\beta)t}\,dt\right\}$$

$$\sim \sum_{m:\,|ma^{-1}-p_\beta|\le za^{-1/2}} \left\{e^{aJ(m/a)}\left(2\pi a^3 r(p_\beta)(1-r(p_\beta))\right)^{-1/2} p_\beta^{-3/2} \times \int_0^{\varepsilon_*} e^{Q(p_\beta)t}\,dt\right\}. \tag{8.5}$$

Step 3. Recall that $J(p)$ attains its maximum value at $p = p_\beta$, and that $J''(p_\beta) < 0$ (cf. (6.9)). Consequently, for m in the range $|ma^{-1} - p_\beta| \le za^{-1/2}$,

$$aJ(m/a) - aJ(p_\beta) = a\left(ma^{-1} - p_\beta\right)^2 J''(p_\beta)/2 + o(1),$$

where $o(1) \to 0$ as $a \to \infty$ uniformly for m in the range $|ma^{-1} - p_\beta| \le za^{-1/2}$. It follows that

$$\sum_{m:\,|ma^{-1}-p_\beta|\le za^{-1/2}} \left(2\pi a^3\right)^{-1/2} e^{aJ(m/a)}$$

$$\sim a^{-1}e^{aJ(p_\beta)} \sum_{m:\,|ma^{-1}-p_\beta|\le za^{-1/2}} (2\pi a)^{-1/2}\exp\left\{a\left(ma^{-1}-p_\beta\right)^2 J''(p_\beta)/2\right\}$$

as $a \to \infty$. Observe that the sum on the right-hand side of this expression is a Riemann sum for $\int_{-z}^{z}(2\pi)^{-1/2}\exp\{y^2J''(p_\beta)/2\}\,dy$. Therefore, as $a \to \infty$,

$$\sum_{m:\,|ma^{-1}-p_\beta|\le za^{-1/2}} \left(2\pi a^3\right)^{-1/2} e^{aJ(m/a)} \sim a^{-1}e^{aJ(p_\beta)}\int_{-z}^{z}(2\pi)^{-1/2}e^{y^2J''(p_\beta)/2}\,dy. \tag{8.6}$$

Notice that as $z \to \infty$ the integral in this expression converges to $(-J''(p_\beta))^{-1/2}$.

A similar argument may be used to bound the sum of those terms in (8.3) in the region $za^{-1/2} < |ma^{-1} - p_\beta| < \delta$. By (8.4), there is a constant $C < \infty$ such that each term is dominated by $Ca^{-3/2}e^{aJ(m/a)}$. If $\delta > 0$ is sufficiently small, Taylor's

theorem implies that

$$aJ(m/a) - aJ(p_\beta) \le a\left(ma^{-1} - p_\beta\right)^2 J''(p_\beta)/4$$

for m satisfying $|ma^{-1} - p_\beta| < \delta$. Hence

$$\sum_{m:\, za^{-1/2} < |ma^{-1} - p_\beta| < \delta} \left(\binom{m}{k_m}\Big/ m + C_m(a;\varepsilon)\right) 1\left\{\left\langle\left\langle (a-m)(\beta-1)^{-1}\right\rangle\right\rangle \ge 1 - \varepsilon_*\right\}$$

$$\le Ca^{-3/2}e^{aJ(p_\beta)} \sum_{m:\, za^{-1/2} < |ma^{-1} - p_\beta| < \delta} \exp\left\{a\left(ma^{-1} - p_\beta\right)^2 J''(p_\beta)/4\right\}$$

$$\le C'a^{-1}e^{aJ(p_\beta)} \int_{|y|>z} e^{y^2 J''(p_\beta)/4}\, dy \tag{8.7}$$

for a suitable constant $C' < \infty$ not depending on a or z. Notice that as $z \to \infty$ the integral in this expression converges to 0.

Combining (8.3), (8.5), (8.6), and (8.7) and letting $z \to \infty$ gives

$$\begin{aligned} N(a;\varepsilon) \sim a^{-1}e^{aJ(p_\beta)}&\left(-J''(p_\beta)\right)^{-1/2}\left(r(p_\beta)\left(1 - r(p_\beta)\right)\right)^{-1/2} p_\beta^{-3/2} \\ &\int_0^{\varepsilon_*} \exp\{Q(p_\beta)t\}\, dt \end{aligned} \tag{8.8}$$

as $a \to \infty$.

9. The "prime number theorem." Define

$$M(a) = \#\{\textit{periodic orbits with period less than } a\}.$$

Then

$$M(a) = \sum_{k \ge 1} N(a - k\varepsilon; \varepsilon) \tag{9.1}$$

for any $\varepsilon > 0$. I shall use the relation (8.8) to obtain an asymptotic formula for $M(a)$ as $a \to \infty$.

It is evident from (8.8) that the largest terms in (9.1) are those for which $a - k\varepsilon$ is large. For these terms the approximation (8.8) is good. For the smaller terms (8.8) may not provide such a good approximation, but these terms do not have much of an effect on the first-order asymptotics of the sum, so there is no harm in using (8.8) to approximate the smaller terms in (9.1) as well. Thus,

$$M(a) \sim C_\varepsilon \sum_{k \ge 1} (a - k\varepsilon)^{-1} e^{(a - k\varepsilon)J(p_\beta)} \tag{9.2}$$

as $a \to \infty$, where

$$C_\varepsilon = \int_0^{\varepsilon_*} e^{Q(p_\beta)t}\, dt \left\{-J''(p_\beta)r(p_\beta)\left(1 - r(p_\beta)\right)\right\}^{-1/2} p_\beta^{-3/2}.$$

Keep in mind that (9.2) is valid for all $\varepsilon > 0$ sufficiently small.

The sum in (9.2) is dominated by those terms for which $(a - k\varepsilon)/a \approx 1$, because of the exponential. Consequently, as $a \to \infty$

$$\begin{aligned} \sum_{k \ge 1} (a - k\varepsilon)^{-1} e^{(a - k\varepsilon)J(p_\beta)} &\sim \sum_{k \ge 1} a^{-1} e^{(a - k\varepsilon)J(p_\beta)} \\ &= a^{-1}e^{aJ(p_\beta)}\left\{e^{-\varepsilon J(p_\beta)}/\left(1 - e^{-\varepsilon J(p_\beta)}\right)\right\}; \end{aligned}$$

hence,

$$M(a) \sim a^{-1} e^{aJ(p_\beta)} C_\varepsilon \{ e^{-\varepsilon J(p_\beta)}/(1 - e^{-\varepsilon J(p_\beta)}) \} \tag{9.3}$$

for all $\varepsilon > 0$. Recall that $\varepsilon_* = \varepsilon(\beta - 1)^{-1}$, so as $\varepsilon \downarrow 0$, $\int_0^{\varepsilon_*} e^{Q(p_\beta)t}\, dt \sim \varepsilon(\beta - 1)^{-1}$. Letting $\varepsilon \downarrow 0$ in (9.3) therefore gives

$$\begin{aligned} M(a) &\sim a^{-1} e^{aJ(p_\beta)} \{ -J''(p_\beta) r(p_\beta)(1 - r(p_\beta)) \}^{-1/2} p_\beta^{-3/2} (\beta - 1)^{-1} J(p_\beta)^{-1} \\ &= (aJ(p_\beta))^{-1} e^{aJ(p_\beta)} \end{aligned} \tag{9.4}$$

(cf. (6.8)–(6.9)). This is the "Prime Number Theorem" for the periodic orbits of the Bernoulli flow.

10. Distribution of individual periodic orbits. The preceding analysis leads to some interesting conclusions about the comportment of "typical" periodic orbits. As we have seen, the sum (5.5) is asymptotically dominated by those terms for which $m/a \sim p_\beta$ and $k_m/m \sim r(p_\beta)$. The index m represents the number of times the periodic orbit crosses the x-axis; the index k_m represents the number of such crossings which occur between $1/2$ and 1 on the x-axis. Hence, "nearly all" periodic orbits with period between a and $a + \varepsilon$ cross the x-axis about ap_β times, and the fraction of such crossings which occur between $1/2$ and 1 is approximately $r(p_\beta)$, provided a is large.

The indices m and k_m are also related to the binary expansions of those points on the x-axis where crossings occur. In particular, m is the period of the binary expansion, while k_m is the number of ones in a single period. Now for most periodic orbits with period between a and $a + \varepsilon$, where a is large, it is the case that $k_m/m \sim r(p_\beta)$. Hence, nearly all periodic orbits cross the x-axis at points whose binary expansions are periodic and have the property that the fraction of ones is about $r(p_\beta)$.

An even stronger statement may be made. Suppose m, k_m are integers satisfying (5.6). Every m-periodic binary expansion with a cycle consisting of k_m ones and $m - k_m$ zeros specifies a periodic orbit of the flow with period between a and $a + \varepsilon$ (except those expansions having period $d|m$, $d \neq m$; however, by (5.2), this complication may be ignored). The collection of such expansions is in 1-1 correspondence with the set of (ordered) random samples (without replacement) of size m from an urn with k_m ones and $m - k_m$ zeros. It is not difficult to show, using the weak law of large numbers and some elementary properties of sampling without replacement, that if m is large and $k_m/m \sim r(p_\beta)$, then "nearly all" such random samples have the following property: the fraction of times the pattern $(e_1, e_2, \dots, e_\lambda)$ of zeros and ones occurs in the ordered random sample is approximately $r(p_\beta)^\nu(1 - r(p_\beta))^{\lambda-\nu}$, where ν is the number of ones in $(e_1, e_2, \dots, e_\lambda)$. (This holds for any fixed pattern $(e_1, e_2, \dots, e_\lambda)$.) Therefore, "nearly all" periodic orbits of the flow with period between a and $a + \varepsilon$, where a is large have the following property: of those points at which the periodic orbit intersects the x-axis, the fraction whose binary expansion begins with the pattern $.e_1e_2 \dots e_\lambda$ is approximately $r(p_\beta)^\nu(1 - r(p_\beta))^{\lambda-\nu}$.

Now if $X_1, X_2, \dots$ is a sequence of independent, Bernoulli-$(r(p_\beta))$ random variables, then the probability that $X_i = e_i$ for each $i = 1, 2, \dots, \lambda$ is also $r(p_\beta)^\nu(1 - r(p_\beta))^{\lambda-\nu}$. Hence, for "nearly all" periodic orbits of the flow with period between a and $a + \varepsilon$, the crossing points of the x-axis have nearly the same

distribution as $X \triangleq .X_1X_2X_3\ldots$. Since each orbit consists of vertical line segments above the crossing points, this specifies (approximately) the distribution of the entire orbit. Recall from section 6 that the maximum entropy distribution for the flow is the distribution of (X, Y), where X is as described above and Y is uniformly distributed on $(0, f_\beta(X))$, with $f_\beta(x) = 1\{x \leq 1/2\} + \beta 1\{x > 1/2\}$. Therefore, *if a is large, then "nearly all" periodic orbits with period between a and $a + \varepsilon$ are approximately distributed according to the maximal entropy measure for the flow.*

REFERENCES

1. R. Bowen, Periodic orbits of hyperbolic flows, *Amer. J. Math.*, 94 (1972) 1–30.
2. ______, Symbolic dynamics for hyperbolic flows, *Amer. J. Math.*, 95 (1973) 429–450.
3. W. Feller, An Introduction to Probability Theory and its Applications, vol. 1, 3rd ed., Wiley, New York, 1968.
4. ______, An Introduction to Probability Theory and its Applications, vol. 2, 2nd ed., Wiley, New York, 1971.
5. H. Furstenberg, Recurrence in Ergodic Theory and Combinatorial Number Theory, Princeton U. Press, Princeton, NJ, 1981.
6. D. Hejhal, The Selberg trace formula and Riemann zeta function, *Duke Math. J.*, 43 (1976) 441–482.
7. S. Lalley, Distribution of periodic orbits of symbolic and Axiom A flows, to appear in *Adv. Appl. Math.* (1987).
8. G. Margulis, Applications of ergodic theory to the investigation of manifolds of negative curvature, *Func. Anal. i Ego Prilozhen*, 3 (1969) 89–90.
9. W. Parry, Bowen's equidistribution theory and the Dirichlet density theorem, Ergodic Th. and Dynam. Syst., 4 (1984) 117–134.
10. W. Parry and M. Pollicott, An analogue of the prime number theorem for closed orbits of Axiom A flows, *Ann. Math.*, 118 (1983) 573–591.
11. P. Sarnak, Asymptotic behavior of the horocycle flow and Eisenstein series, *Comm. Pure and Applied Math.*, 34 (1981) 719–739.
12. C. Series, Symbolic dynamics for geodesic flows, *Acta. Math.*, 146 (1981) 103–128.

Folding Polynomials and Their Dynamics

WM DOUGLAS WITHERS, *U.S. Naval Academy*

WM DOUGLAS WITHERS: I received my Ph.D. in 1983 from Georgia Tech, working under Les Karlovitz. I spent a year at the University of Maryland as a visitor before coming to the Naval Academy in 1984. Most of my research has been in the somewhat related fields of dynamical systems and analysis of fractals. In the past year I have become interested also in the theory of orthogonal polynomials in several variables. One might describe this as entirely disjoint from my other interests, except that this article serves as a counterexample.

Abstract. The equilateral triangle, the right isoceles triangle, and the 30–60–90 triangle are special in that they can be folded into replicas of themselves. We describe polynomial mappings which are equivalent as dynamical systems to the mappings which stretch and fold these triangles onto themselves. The analogous polynomials in one dimension are the Chebyshev polynomials, whose special properties as dynamical systems are shared by our polynomials. Our construction can be carried out in any number of dimensions.

1. Introduction. The mapping

$$f(x) = 2x^2 - 1 \tag{1.1}$$

on the interval $[-1, 1]$ is one of the most basic examples in the theory of dynamical systems. The dynamical systems view of a mapping such as (1.1) is to consider the points in $[-1, 1]$ as possible states of a physical system and the mapping $x \mapsto f(x)$ as describing the evolution of the system over one unit of time. It is in this sense that we refer to a mapping such as f from a set into itself as a *dynamical system*. The long-term behavior of the physical system is described by the behavior of the iterates x_0, $x_1 = f(x_0)$, $x_2 = f(x_1), \ldots, x_{n+1} = f(x_n), \ldots$ as n goes to infinity. We call the set of these iterates the *orbit* of x_0.

Typical questions to be asked about a dynamical system such as f are:

Does f have periodic orbits of a given period p? A *periodic* orbit of period p is a set of p points $\{x_1, \ldots, x_p\}$ such that $f(x_n) = x_{n+1}$ for $n = 1, \ldots, p-1$ and $f(x_p) = x_1$. In particular, a fixed point for f is a periodic orbit of period 1.

Is there a measure μ which gives long-term average values for continuous functions g:

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} g(f^n(x_0)) = \int g \, d\mu? \tag{1.2}$$

Such a measure must be invariant under f; that is, $\mu E = \mu f^{-1}(E)$ for measurable sets E, and often μ is largely independent of x_0.

Do iterates of initial conditions close together separate over time and if so, how rapidly? Relevant to this question is the determination of *Lyapunov exponents*, logarithms of average stretching factors. For example, a one-dimensional mapping f such as (1.1) has a single Lyapunov exponent λ given by

$$\lambda = \lim_{k \to \infty} \log|D(f^k)(x)|^{1/k},$$

where D represents the derivative. Often λ is largely independent of x.

We shall use the term *dynamical properties* loosely to refer to properties such as the existence of periodic orbits, invariant measures, and Lyapunov exponents.

By an *attractor* for f we mean a compact set A such that the set

$$B = \left\{x\colon \lim_{n\to\infty} \inf_{y\in A} \|f^n(x) - y\| = 0\right\},$$

has positive Lebesgue measure, and which is minimal with respect to this property. The set B is called the *basin of attraction* for A. One can show that A and B must be invariant under f: $f(A) = A$, $f(B) = B$. If a measure μ giving long-term averages as in (1.2) is independent of x_0 for x_0 in some set E of positive Lebesgue measure, then its support A is an attractor for the system and E is contained in the basin of attraction for A.

An attractor may be of classical type, a periodic orbit or (in higher-dimensional spaces) a torus. For the mapping (1.1), however, it can be shown that the iterates of almost any initial condition in $[-1,1]$ are dense in $[-1,1]$; this interval is thus an attractor for f, a *strange* attractor, as attractors not of classical type are called.

The system (1.1) is, however, not at all typical of systems with strange attractors because its dynamical properties can be determined in great detail. Although for most systems with strange attractors, the problems of counting periodic orbits, representing invariant measures, calculating Lyapunov exponents, and so forth, seem quite intractable, for (1.1) all these questions can be answered with the aid of the coordinate change $x = h(u) = \cos \pi u$. When we make this change of coordinates, the system is transformed to the "tent map":

$$t(u) = h^{-1} \circ f \circ h(u) = \begin{cases} 2u & \text{if } u \le \dfrac{1}{2}, \\ 2(1-u) & \text{if } u \ge \dfrac{1}{2}, \end{cases}$$

on $[0,1]$. Since this mapping is piecewise linear with a constant stretching factor 2, we can see at once that its Lyapunov exponent is $\log 2$. We can determine other dynamical properties with the aid of a symbolic argument. As applied to the tent map, the symbolic argument consists mainly of writing u as a binary expansion and noting that the map t merely shifts and (sometimes) complements the binary expansion of u. We present a more detailed example of such an argument in §5 of this article. Almost every point in $[0,1]$ has a binary expansion which is statistically random, that is, in which all sequences of digits of a certain finite length occur with equal frequency. It follows that the iterates $t^n(u)$ of almost every initial condition in $[0,1]$ are uniformly distributed in $[0,1]$ as n goes to infinity; thus the invariant measure which gives long-term average values is just Lebesgue measure on $[0,1]$. We can also count periodic orbits of t by observing that the points of these orbits are just those points with periodic binary expansions.

Most dynamical properties are easily transformed under a differentiable change of coordinates; thus we can determine the properties of our quadratic mapping f. The Lyapunov exponent is invariant under such a coordinate change, so the Lyapunov exponent of f is $\log 2$. The invariant measure μ giving long-term average values for f is the image of Lebesgue measure under the cosine function; the

measure of a set E is therefore given by

$$\mu E = \int_E \frac{dx}{\pi\sqrt{(1-x^2)}}.$$

Periodic orbits of t map directly to periodic orbits of f under the cosine function.

If we seek additional examples of dynamical systems with strange attractors whose dynamical properties can nonetheless be completely determined, we observe that the Chebyshev polynomials, by virtue of the definition

$$T_n(\cos \pi u) = \cos n\pi u \tag{1.3}$$

all share the properties of (1.1); under the transformation $x = \cos \pi u$, the nth Chebyshev polynomial T_n becomes a piecewise linear function which stretches the interval $[0, 1]$ by a factor of n and folds it back n times onto itself. Examples of such completely solvable systems beyond this seem to be scarce.

The purpose of this article is to show that further examples do exist and have many interesting properties. For example, the mapping

$$A_2(x) = x^2 - 2\bar{z}$$

in the complex plane has as an attractor a deltoid with vertices at the points 3 and $-3/2 \pm 3\sqrt{3}/2$ with its interior, as shown in Figure 1.

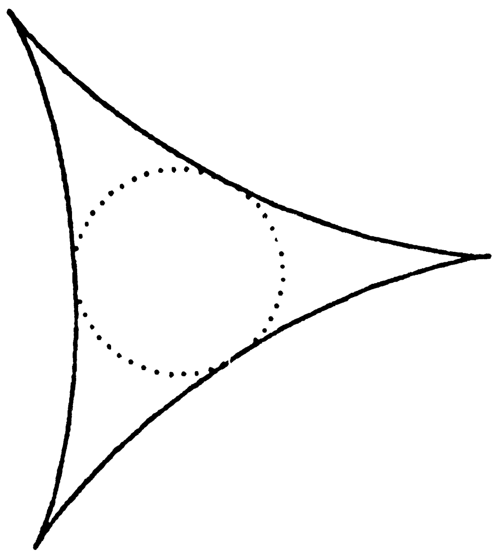


FIG. 1.

The dotted circle with its interior is mapped under A_2 to the entire attractor, as are each of the corner regions. This system can be analysed with the aid of a certain coordinate change (detailed in §3) which transforms the deltoid into an equilateral triangle and its inscribed circle into an inscribed triangle, as in Figure 2. Under this change of coordinates, the mapping A_2 from the deltoid to itself becomes a

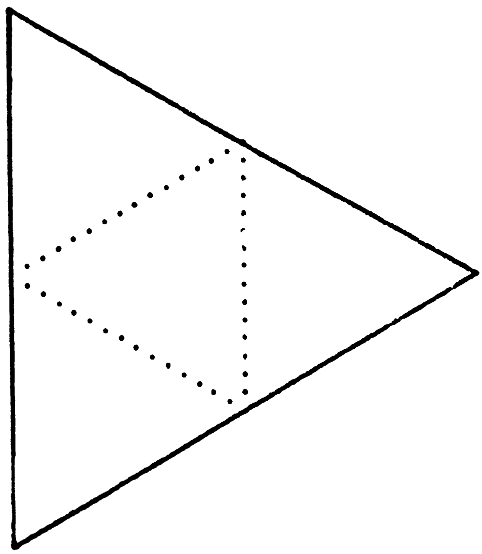


FIG. 2.

mapping F_2 from the triangle to itself, and F_2 is nothing more than folding the triangle on the dotted lines and stretching by a factor of two to cover the original triangle (the vertex at the right is mapped to itself). The mapping A_2 is one of a sequence of polynomial mappings which are equivalent to folding the triangle into $9, 16, 25, \ldots$ copies of itself.

By a *foldable* shape we mean a polytope which can be folded into replicas of itself. Corresponding to any foldable shape in k dimensions there is a sequence of polynomials P_n such that the mapping P_n is equivalent to stretching and folding the shape onto itself n^k-to-one. We call these polynomials *folding polynomials*. More precisely, a foldable shape Φ is one which can be partitioned into smaller congruent replicas $\phi_1, \ldots, \phi_{n^k}$ of itself which overlap only at their boundaries, and for which there exists a mapping F_n: $\Phi \to \Phi$ which is continuous and whose restriction to ϕ_j is n times an isometry. In fact if F_n exists for any n then it exists for all natural numbers n. We can uniquely define this mapping by requiring that it be n times the identity in the neighborhood of a particular point of Φ (in two dimensions, the vertex with the smallest angle). Then there exists a function h defined on Φ and a polynomial P_n which is conjugate to F_n by h; that is, P_n: $h(\Phi) \to h(\Phi)$ and

$$P_n = h \circ F_n \circ h^{-1}.$$

This existence theorem is proven in Hoffman and Withers [3].

In two dimensions there are just four foldable shapes: the equilateral triangle, the isoceles right triangle, the 30–60–90 triangle, and the rectangle, as is shown in [3]. Since the rectangle is a product of intervals, it is a trivial matter to define the associated polynomials in that case in terms of Chebyshev polynomials. Associated with each of the other three shapes is a sequence of folding polynomials, which can be transformed by a suitable change of coordinates into stretching and folding of the corresponding triangle onto itself.

We have not explored applications of folding polynomials in great depth; we note, however, that the polynomials we call A_n were used by Mittag and Stephen [**6**] in the study of the Potts model of phase transitions. They are also related to orthogonal polynomials in several variables studied by Koornwinder [**4**] and Eier and Lidl [**2**].

A complete description of the derivation and theory of folding polynomials is given in [**3**]. In this article we sketch the derivation and properties of the particular sequence of polynomials A_n associated with the equilateral triangle and discuss their properties as dynamical systems, together with other special properties. We then present briefly the sequences B_n and G_n associated with the right isoceles triangle and the 30–60–90 triangle, respectively. Proofs are omitted where our results can be verified by direct computation.

2. The (u, v) plane. We start with two lines in the plane $\mathbf{R}^2$, denoted l_u and l_v, which intersect at an angle of $\pi/3$. We give each point in the plane coordinates (u, v), where u is the signed distance of the point from l_u and v its signed distance from l_v. We choose signs so that the region where both u and v are positive has an angle of $\pi/3$. The region bounded by the lines $v = -2u$, $u = -2v$, and $u - v = 1$ is an equilateral triangle, which we shall call the *fundamental region* and denote by Φ; in general, we denote by Φ_n the triangle bounded by the lines $v = -2u$, $u = -2v$, and $u - v = n$. This system of coordinates is shown in Figure 3.

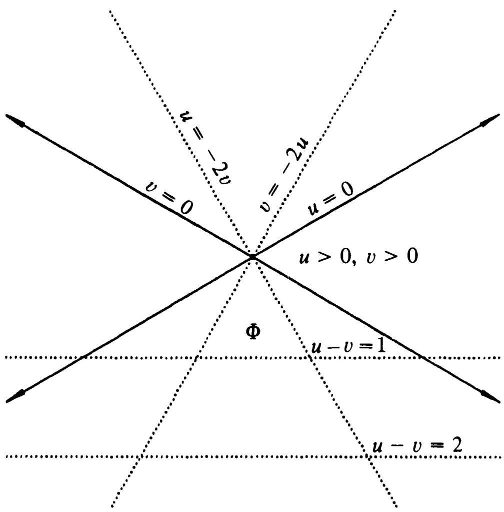


FIG. 3.

3. The function h and its properties. Let h: $\mathbf{R}^2 \to \mathbf{C}$ be defined by

$$h(u, v) = e^{2\pi i u} + e^{2\pi i v} + e^{-2\pi i(v+u)}.$$

Let ρ_1: $\mathbf{R}^2 \to \mathbf{R}^2$ be defined by $\rho_1(u, v) = (1 + v, u - 1)$; then ρ_1 is a reflection through the line $u - v = 1$. Let ρ_2: $\mathbf{R}^2 \to \mathbf{R}^2$ be defined by $\rho_2(u, v) = (u, -u - v)$; then ρ_2 is a reflection through the line $u = -2v$. Let ρ_3: $\mathbf{R}^2 \to \mathbf{R}^2$ be defined by

$\rho_3(u, v) = (-v -u, v)$; then ρ_3 is a reflection through the line $v = -2u$. We then have

$$h \circ \rho_1 = h \circ \rho_2 = h \circ \rho_3 = h.$$

Thus h is invariant under any isometry in the group $\mathbf{A}$ generated by ρ_1, ρ_2, and ρ_3. It follows that the values taken by h everywhere in the plane are determined by its values in Φ.

Since h is continuous, it must map Φ and hence the entire plane onto a compact subset of $\mathbf{C}$. We denote $h(\mathbf{R}^2)$ by Δ. The boundary $\partial\Delta$ of Δ is the image under h of the set of points where the Jacobian of h is zero; this set consists of the line $v = -2u$ and its images under the group $\mathbf{A}$. We can thus find by substitution:

$$\partial\Delta = \{h(u, -2u): u \in \mathbf{R}\} = \{2e^{2\pi iu} + e^{-4\pi iu}: u \in \mathbf{R}\}.$$

This parametric equation describes a hypocycloid of three cusps, called a deltoid. Some properties of this curve are summarized in [5]. The set Δ consists of the deltoid with its interior as in Figure 1. The function h is a bijection between the boundary of Φ and $\partial\Delta$ and between Φ and Δ. The vertices of Φ are mapped to the cusps of Δ.

Let us also note how h^{-1} may be calculated. If we let $z = h(u, v)$, note that we have

$$e^{2\pi iu}e^{2\pi iv} + e^{2\pi iv}e^{-2\pi i(v+u)} + e^{-2\pi i(v+u)}e^{2\pi iu} = \bar{z}$$

and

$$e^{2\pi iu}e^{2\pi iv}e^{-2\pi i(v+u)} = 1.$$

Thus the roots of the equation

$$s^3 - zs^2 + \bar{z}s - 1 = 0 \tag{3.1}$$

are $e^{2\pi iu}$, $e^{2\pi iv}$, and $e^{-2\pi i(v+u)}$. This shows also that h is one-to-one from Φ to Δ.

4. Definition of A_n; the recursion formula. Let us define F_n, the nth "folding function", from Φ to Φ_n as follows. Define S_n: $\Phi \to \Phi_n$ by $S_n(u, v) = (nu, nv)$. The triangle Φ_n consists of n^2 copies $\Psi_1, \ldots, \Psi_{n^2}$ of the triangle Φ; for each copy Ψ_k there is an element α_k of $\mathbf{A}$ which maps it onto Φ. We define U_n: $\Phi_n \to \Phi$ by $U_n(u, v) = \alpha_k(u, v)$ for $(u, v) \in \Psi_k$. Then $F_n = U_n \circ S_n$ maps Φ n^2-to-one onto itself, except for points mapped to the boundary.

We define the nth folding polynomial A_n: $\Delta \to \Delta$ by

$$A_n = h \circ F_n \circ h^{-1}. \tag{4.1}$$

Since $h \circ U_n = h$, we have the alternative definition

$$A_n(h(u, v)) = h(nu, nv). \tag{4.2}$$

Note the similarity to the definition (1.3) of the Chebyshev polynomials.

The functions defined in this way are nonanalytic functions of z but are polynomials in z and $\bar{z}$. To see this, one may verify

$$A_0(z) = 3.$$

$$A_1(z) = z.$$

$$A_2(z) = z^2 - 2\bar{z}.$$

That A_n is indeed a polynomial in z and $\bar{z}$ for all n then follows from the recursion formula:

$$A_n = zA_{n-1} - \bar{z}A_{n-2} + A_{n-3} \qquad (n \geq 3).$$

Thus $A_n(z)$ can be defined for all $z \in C$. It is a simple matter to calculate further A_n:

$$\begin{aligned} A_3(z) &= z^3 - 3z\bar{z} + 3, \\ A_4(z) &= z^4 - 4z^2\bar{z} + 2\bar{z}^2 + 4z, \\ A_5(z) &= z^5 - 5z^3\bar{z} + 5z\bar{z}^2 + 5z^2 - 5\bar{z}, \end{aligned}$$

and so forth.

From the definition (4.1) we can obtain directly the fact that, for $n > 0$, A_n maps Δ onto itself n^2-to-one, as well as the composition property:

$$A_{mn} = A_m \circ A_n = A_n \circ A_m.$$

The composition property is an unusual one; it is known, for example, that the only sequences of one-dimensional polynomials with this property are essentially the sequence of Chebyshev polynomials $T_n(x)$ and the sequence of power functions x^n [7].

5. Geometric mapping properties of the A_n. One property of the sequence of power functions $z \mapsto z^n$ on $\mathbf{C}$ is that they carry certain straight lines into straight lines, namely, those lines containing 0. The functions A_n have a similar property, but for a different family of lines.

Consider a line in the (u, v) plane given by the equation $u = c$, c being constant. Its image L_c under h is a line with parametric equation

$$\begin{aligned} z &= e^{2\pi i c} + e^{2\pi i v} + e^{-2\pi i(v+c)} \\ &= e^{2\pi i c} + e^{-2\pi i c/2}\left(e^{2\pi i(v+c/2)} + e^{-2\pi i(v+c/2)}\right). \end{aligned}$$

The quantity in parentheses can be considered as a real parameter

$$t = 2\cos 2\pi(v + c/2);$$

thus the equation becomes

$$z = e^{2\pi i c} + e^{-\pi i c}t.$$

The line L_c makes an angle of $c/2$ with the positive real axis.

From equation (4.2), the image of L_c under the map A_n is the line L_{nc}. Thus, for the family of lines L_c, the functions A_n have the same angle multiplication property as the power functions $z \mapsto z^n$ have for lines passing through the origin.

In Figure 4 we have plotted the lines L_c for several values of c. It is seen that their envelope is the deltoid $\partial\Delta$.

Directing our attention to the function A_2, let us inquire where its Jacobian is zero. Since $A_2(z) = z^2 - 2\bar{z}$, the derivative matrix for A_2 is

$$\begin{pmatrix} 2x - 2 & -2y \\ 2y & 2x + 2 \end{pmatrix},$$

where $z = x + iy$. Thus the set where the Jacobian of A_2 is zero is a unit circle centered at the origin; this is the inscribed circle Γ of the deltoid Δ. If we

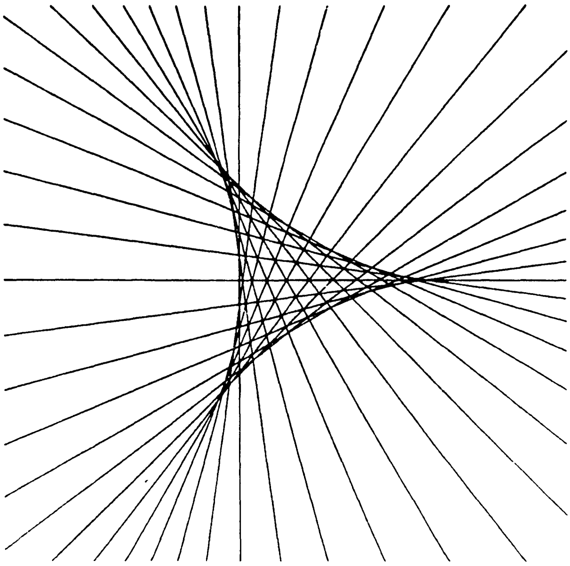


FIG. 4.

parametrize this circle by $z = e^{i\theta}$, we find that its image under the map A_2 is the curve $z = e^{2i\theta} + e^{-i\theta}$, which is $\partial\Delta$.

6. Dynamical properties of the A_n. The fact that the polynomials A_n on the deltoid Δ can be transformed by the differentiable coordinate change h to the folding functions F_n on the triangle Φ makes it possible to determine their dynamical properties, that is, the behavior of the iterates $A_n(z), A_n^2(z), \ldots$ of a point z in great detail. One must first determine the dynamical properties of the functions F_n. Since the function F_n is piecewise linear, with a derivative matrix which is n times an isometry, this can be accomplished with the aid of a symbolic argument, which will be familiar to dynamicists.

Subdivide Φ into equilateral triangles $\phi_1, \ldots, \phi_{n^2}$, each of which is $1/n$ the scale of Φ and is mapped onto Φ by F_n. These triangles overlap only on their boundaries. We further divide each ϕ_j into smaller equilateral triangles $\phi_{j1}, \ldots, \phi_{jn^2}$ and each ϕ_{jk} into smaller triangles $\phi_{jk1}, \ldots, \phi_{jkn^2}$, and so on, such that $\phi_{j_0 \cdots j_{k+1}}$ is $1/n$ the scale of $\phi_{j_0 \cdots j_k}$ and $F_n(\phi_{j_0 \cdots j_{k+1}}) = \phi_{j_1 \cdots j_{k+1}}$. Thus $\operatorname{diam} \phi_{j_0 \cdots j_k} = n^{-k-1} \operatorname{diam} \Phi$ and a triangle $\phi_{j_0 \cdots j_k}$ overlaps others of the same size only at its boundary.

Let Σ be the set of infinite sequences from $\{1, \ldots, n^2\}$. Corresponding to each $(j_0, j_1, \ldots) \in \Sigma$ is a unique point $\phi(j_0, j_1, \ldots) \in \Phi$ given by

$$\{\phi(j_0, j_1, \ldots)\} = \bigcap_{k=0}^{\infty} \phi_{j_0 \cdots j_k}.$$

Moreover, we have

$$F_n \circ \phi = \phi \circ \sigma,$$

where $\sigma \colon \Sigma \to \Sigma$ is the left shift operator:

$$\sigma(j_0, j_1, \ldots) = (j_1, j_2, \ldots).$$

We also have the relationship $F_n^k(\phi(j_0, j_1, \ldots)) \in \phi_{j_k}$. Thus a point (u, v) of Φ which corresponds to more than one sequence in Σ must lie on the boundary of some $\phi_{j_0 \cdots j_k}$ and $F_n^{k+1}(u, v)$ lies on the boundary of Φ. From this it can be shown that the set of points in Φ which correspond to more than one sequence has Lebesgue area measure zero. We can now establish the following dynamical properties of F_n.

(i) Periodic orbits of F_n can be counted as follows. It can be shown that each point of a periodic orbit of F_n of period p corresponds to a unique sequence in Σ. This sequence must therefore be a point of a periodic orbit for σ of period p; in other words, the sequence must be periodic. We thus have the combinatorial problem of counting the periodic sequences with a given period. This is known as the circular word problem and its solution is the following. Let $\Pi(p, n)$ denote the number of different circular words of p letters on an alphabet of n^2 letters. Then (as in Berge [**1**])

$$\Pi(p, n) = \frac{1}{p} \sum_{q|p} \mu\left(\frac{p}{q}\right) n^{2q}, \tag{6.1}$$

where μ is the Möbius function ($\mu(s) = (-1)^j$ if s is a product of j distinct primes, $\mu(s) = 0$ otherwise) and the symbol | denotes "divides."

(ii) The Lyapunov exponents of F_n are both equal to $\log n$. (A mapping f: $\mathbf{R}^2 \to \mathbf{R}^2$ has two Lyapunov exponents $\lambda_1 \geq \lambda_2$ given by

$$\lambda_1 = \lim_{k \to \infty} \log \| D(f^k)(u, v) \|^{1/k},$$

$$\lambda_2 = \lim_{k \to \infty} -\log \| D(f^k)(u, v)^{-1} \|^{1/k},$$

where $Df(u, v)$ denotes the derivative matrix of f at (u, v).) In the case of F_n these limits take the same value, $\log n$, unless (u, v) is in the set of measure zero which is eventually mapped to the boundary of Φ, some power of F_n then being nondifferentiable at (u, v).

(iii) Almost all initial conditions in Φ have orbits under F_n which are dense in Φ, and the invariant measure giving long-term average values of continuous functions is uniform on the triangle Φ. Since the area of the triangle is $1/6$, the density is 6. To see this, we choose a point (u, v) at random from Φ, uniformly distributed, and let $(j_0, j_1, \ldots)$ be its corresponding sequence in Σ (unique with probability 1). The probability that j_k takes a particular value l is given by $m\phi_{j_0 \cdots j_{k-1}l}/m\phi_{j_0 \cdots j_{k-1}} = 1/n^2$, where m is Lebesgue area measure. Thus, with probability one, any finite sequence $(l_1, \ldots, l_k)$ appears in the infinite sequence $(j_0, j_1, \ldots)$ with frequency $m\phi_{l_1 \cdots l_k}/m\Phi = 1/n^{2k}$, and the orbit of (u, v) visits $\phi_{l_1 \cdots l_k}$ with this frequency. From this it can be shown that the orbit of (u, v) visits any measurable set E with frequency $mE/m\Phi = mE/6$ with probability one.

The corresponding dynamical properties of the functions A_n can be derived from these. Properties (i) and (ii) above are invariant under a differentiable coordinate change such as h. Thus we have

(i) The function A_n has $\Pi(p, n)$ periodic orbits of period p, where Π is as in equation (6.1).

(ii) The Lyapunov exponents of A_n are both equal to $\log n$.

The invariant measure generated by A_n is the image of that for F_n under h. Thus it is absolutely continuous and has density

$$\delta(z) = \frac{6}{J(h) \circ h^{-1}},$$

where $J(h)$ denotes the Jacobian determinant of h. Thus δ satisfies

$$\delta(h(u, v)) = \frac{6}{J(h)(u, v)}. \tag{6.2}$$

Note that $J(h)$ cannot be a function of $z = h(u, v)$ because it is not invariant under the reflections ρ_1, ρ_2, and ρ_3; it actually changes sign under each of these reflections. However, it is true that $(J(h))^2$ is invariant under ρ_1, ρ_2, and ρ_3 and is a function of z, indeed a polynomial in z and $\bar{z}$. It is a result of the theory of folding polynomials that for any folding polynomial, the density of the invariant measure is the inverse square root of a polynomial.

Theoretically one could determine h^{-1} from equation (3.1) and compose it with $J(h)$ to determine δ. What we did in fact was to take the equation of the deltoid in Cartesian coordinates from [**5**]:

$$(x^2 + y^2)^2 - 8x(x^2 - 3y^2) + 18(x^2 + y^2) - 27 = 0.$$

In terms of $z = x + iy$ this becomes

$$\sigma(z) = (z\bar{z})^2 - 4(z^3 + \bar{z}^3) + 18z\bar{z} - 27 = 0.$$

Since $J(h)$ is zero on the boundary of Φ, the density δ goes to infinity on the boundary of Δ. It is not unreasonable to suppose that δ might be somehow inversely related to σ. With some experimentation we arrived at the following formula which satisfies (6.2):

$$\delta(z) = \frac{3}{\pi^2}\left[27 - 18z\bar{z} - (z\bar{z})^2 + 4(z^3 + \bar{z}^3)\right]^{-1/2}$$

The polynomials A_n are also orthogonal with respect to the density δ; it is proven in [**3**] that:

$$\int_\Delta A_m(x + iy) \cdot \overline{A_n(x + iy)}\delta(x + iy)\, dx\, dy = 0.$$

The sequence A_n is not a complete set; one can define, however, a complete set of polynomials associated with the A_n in a natural way; see again [**3**].

7. Polynomials B_n associated with the right isoceles triangle. In the case of the right isoceles triangle we will use a standard coordinate system in the (u, v) plane with the axes intersecting at right angles. Our fundamental region Φ_B consists of the right isoceles triangle bounded by the lines $u = 0$, $v = 1/2$, and $u = v$.

The crucial factor in finding a sequence of polynomials conjugate to stretching the triangle Φ_B and folding it onto itself is to find a function h_B such that (i) h_B is invariant under the group of isometries generated by the reflections through the edges of Φ_B and (ii) $h_B(nu, nv)$ is a polynomial in $h_B(u, v)$. We define h_B: $\mathbf{R}^2 \to \mathbf{R}^2$ by

$$h_B(u, v) = (2\cos 2\pi u + 2\cos 2\pi v, 4\cos 2\pi u \cos 2\pi v).$$

Let $\rho_1(u, v) = (-u, v)$ be the reflection through the line $u = 0$. Let $\rho_2(u, v) = (u,$

$1 - v)$ be the reflection through the line $v = 1/2$. Let $\rho_3(u, v) = (v, u)$ be the reflection through the line $u = v$. Then we have

$$h_B \circ \rho_1 = h_B \circ \rho_2 = h_B \circ \rho_3 = h_B,$$

so h_B is invariant under any isometry in the group $\mathbf{B}$ generated by ρ_1, ρ_2, and ρ_3, and the values taken by h_B everywhere in the plane are determined by its values in Φ_B.

The function h_B maps the entire plane onto a compact subset of $\mathbf{R}^2$, which we denote Δ_B. The boundary of Δ_B is the image under h_B of the lines $u = 0$, $v = 1/2$, and $u = v$ which bound Φ_B. One can show that these images are the lines $y + 4 - 2x = 0$, $y + 4 + 2x = 0$, and the parabola $x^2 = 4y$, respectively; these form the boundary of Δ_B. Figure 5 shows this region.

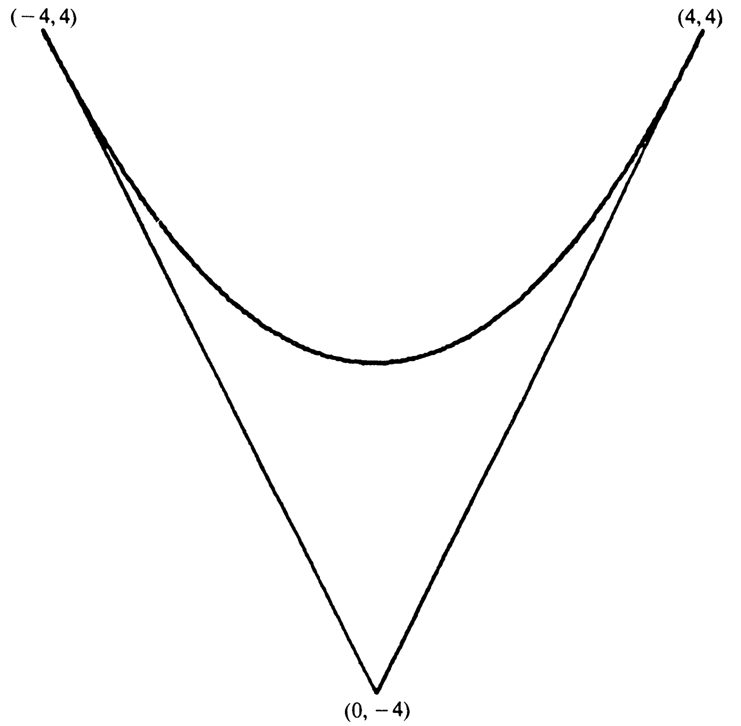


FIG. 5.

Parallel to the definition (4.2) of the polynomials A_n we define B_n: $\Delta_B \to \Delta_B$ by

$$B_n\big(h_B(u, v)\big) = h_B(nu, nv). \tag{7.1}$$

We then have

$$\begin{aligned}
B_0(x, y) &= (4, 4)\\
B_1(x, y) &= (x, y)\\
B_2(x, y) &= \left(x^2 - 2y - 4,\ y^2 - 2x^2 + 4y + 4\right)\\
B_3(x, y) &= \left(x^3 - 3xy - 3x,\ y^3 - 3x^2y + 6y^2 + 9y\right).
\end{aligned}$$

That B_n is indeed a polynomial for all n then follows from the recursion formulae:

$$x_{n+4} = x(x_{n+3} + x_{n+1}) - (2 + y)x_{n+2} - x_n, \tag{7.2}$$

$$y_{n+4} = y(y_{n+3} + y_{n+1}) - (x^2 - 2y - 2)y_{n+2} - y_n, \tag{7.3}$$

where we have written (x_n, y_n) for $B_n(x, y)$. It is then a straightforward matter to calculate further B_n:

$$\begin{aligned} B_4(x, y) = \big(&x^4 - 4x^2y - 4x^2 + 2y^2 + 8y + 4,\\ &y^4 - 4x^2y^2 + 2x^4 + 8y^3 - 8x^2y + 20y^2 - 8x^2 + 16y + 4\big),\\ B_5(x, y) = \big(&x^5 - 5x^3y - 5x^3 + 5xy^2 + 15xy + 5x,\\ &y^5 - 5x^2y^3 + 5x^4y + 10y^4 - 20x^2y^2\\ &+ 35y^3 - 25x^2y + 50y^2 + 25y\big), \end{aligned}$$

and so forth.

From the definition (7.1) we can obtain directly the fact that, for $n > 0$, B_n maps Δ_B onto itself n^2-to-one, as well as the composition property:

$$B_{mn} = B_m \circ B_n = B_n \circ B_m. \tag{7.4}$$

The right isoceles triangle differs from the equilateral triangle in that besides being foldable into n^2 copies of itself it can be folded into 2 (and therefore $2n^2$) copies of itself. It is natural to ask whether there also exist polynomials conjugate to these mappings. Such polynomials do exist; they are given by the definition:

$$B_{n\sqrt{2}}(h_B(u, v)) = h_B(nu + nv, nu - nv). \tag{7.5}$$

For example we have

$$B_{\sqrt{2}}(x, y) = (y, x^2 - 2y - 4). \tag{7.6}$$

With definition (7.5), the composition formula (7.4) holds even if m or n or both are multiples of $\sqrt{2}$. Thus the recursion formulae (7.2) and (7.3) together with (7.6) allow us to calculate all of this "extra" sequence of polynomials.

The dynamical properties of the polynomials B_n can be obtained in the same way as those of the A_n; we list them here (n represents an integer or an integral multiple of $\sqrt{2}$):

(i) The function B_n has $\Pi(p, n)$ periodic orbits of period p in Δ_B, where Π is as in equation (6.1).

(ii) The Lyapunov exponents of B_n are both equal to $\log n$.

(iii) Almost all initial conditions of Δ_B have orbits which are dense in Δ_B, and the invariant measure generated by B_n is absolutely continuous with density

$$\delta(x, y) = \frac{2}{\pi^2}\left[(y + 4 + 2x)(y + 4 - 2x)(x^2 - 4y)\right]^{-1/2}.$$

The polynomials B_n are also orthogonal with respect to this measure.

8. Polynomials G_n associated with the 30–60–90 triangle. For the 30–60–90 triangle we use the same coordinate system in the (u, v) plane as for the equilateral triangle; the axes intersect at an angle of $\pi/3$. Our fundamental region Φ_G consists of the 30–60–90 triangle bounded by the lines $u = -2v$, $u = -v$, and $u - v = 1$.

We then define $h_G\colon \mathbf{R}^2 \to \mathbf{R}^2$ by

$$h_G(u,v) = (2\cos 2\pi u + 2\cos 2\pi v + 2\cos 2\pi(v+u),$$
$$2\cos 2\pi(u-v) + 2\cos 2\pi(u+2v) + 2\cos 2\pi(v+2u)).$$

Let $\rho_1(u,v) = (u, -u-v)$ be the reflection through the line $u = -2v$. Let $\rho_2(u,v) = (-v,-u)$ be the reflection through the line $u = -v$. Let $\rho_3(u,v) = (1+v, u-1)$ be the reflection through the line $u - v = 1$. Then we have

$$h_G \circ \rho_1 = h_G \circ \rho_2 = h_G \circ \rho_3 = h_G,$$

so h_G is invariant under any isometry in the group $\mathbf{G}$ generated by ρ_1, ρ_2, and ρ_3, and the values taken by h_G everywhere in the plane are determined by its values in Φ_G.

The function h_G maps the entire plane onto a compact subset of $\mathbf{R}^2$, which we denote Δ_G. The boundary of Δ_G is the image under h_G of the lines $u = 2v$, $u = v$, and $u + v = 1$ which bound Φ_G. These images are curves described by parametric equations

$$\begin{cases} x = 4t^2 + 4t - 2, \\ y = 16t^3 - 12t + 2, \end{cases}$$

and

$$\begin{cases} x = 2 + 4t, \\ y = 4t^2 + 4t - 2. \end{cases}$$

(The lines $u = 2v$ and $u + v = 1$ map into the same curve.) These curves form the boundary of Δ_G. Figure 6 shows this region.

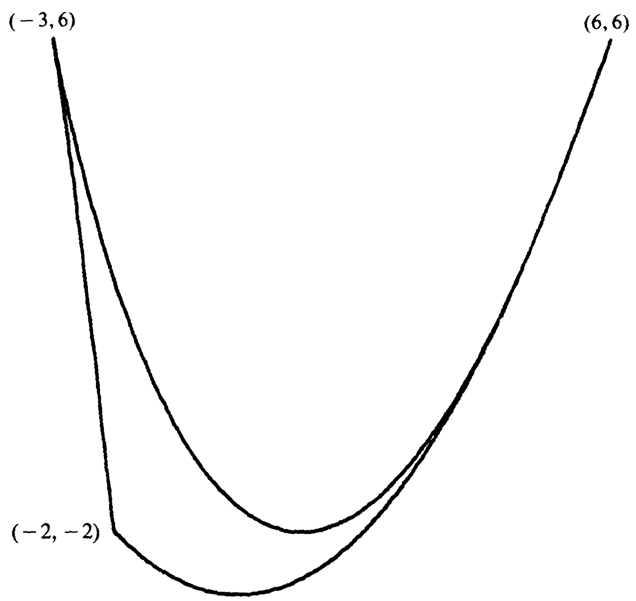


FIG. 6.

Parallel to the definition (4.2) of the polynomials A_n we define G_n: $\Delta_G \to \Delta_G$ by

$$G_n(h_G(u,v)) = h_G(nu, nv). \tag{8.1}$$

We then have

$$G_0(x,y) = (6,6).$$

$$G_1(x,y) = (x,y).$$

$$G_2(x,y) = (x^2 - 2x - 2y - 6,\ y^2 - 2x^3 + 6xy + 10y + 18x + 18).$$

$$\begin{aligned}G_3(x,y) = (&x^3 - 3xy - 9x - 6y - 12,\ y^3 - 3x^3y + 9xy^2\\ &-6x^3 + 18y^2 + 45xy + 63y + 54x + 60).\end{aligned}$$

$$\begin{aligned}G_4(x,y) = (&x^4 - 4x^2y - 10x^2 - 4xy + 2y^2 - 8x + 8y + 6,\\ &y^4 + 2x^6 - 4x^3y^2 - 12x^4y + 12xy^3 + 18x^2y^2\\ &-28x^3y - 36x^4 + 24y^3 + 120xy^2\\ &+108x^2y - 40x^3 + 134y^2 + 372xy + 162x^2 + 280y + 360x + 198).\end{aligned}$$

$$\begin{aligned}G_5(x,y) = (&x^5 - 5x^3y - 15x^3 - 5x^2y + 5xy^2 - 10x^2\\ &+35xy + 10y^2 + 55x + 50y + 60,\\ &y^5 + 5x^6y - 5x^3y^3 - 30x^4y^2 + 10x^6 + 15xy^4\\ &+45x^2y^3 - 65x^3y^2 - 150x^4y\\ &+30y^4 + 240xy^3 + 360x^2y^2 - 205x^3y - 180x^4 + 255y^3 + 1200xy^2\\ &+945x^2y - 190x^3 + 920y^2 + 2415xy + 810x^2 + 1495y + 1710x + 900).\end{aligned}$$

That G_n is indeed a polynomial for all n then follows from the recursion formulae:

$$\begin{aligned}x_{n+6} &= x(x_{n+5} + x_{n+1}) - (x + y + 3)(x_{n+4} + x_{n+2})\\ &\quad + (x^2 - 2y - 4)x_{n+3} - x_n,\end{aligned} \tag{8.2}$$

$$\begin{aligned}y_{n+6} &= y(y_{n+5} + y_{n+1}) - (x^3 - 3xy - 9x - 5y - 9)(y_{n+4} + y_{n+2})\\ &\quad + (y^2 - 2x^3 + 6xy + 18x + 12y + 8)y_{n+3} - y_n,\end{aligned} \tag{8.3}$$

where we have written (x_n, y_n) for $G_n(x,y)$. It is then a straightforward matter to calculate further G_n.

From the definition (8.1) we can obtain directly the fact that G_n maps Δ_G onto itself n^2-to-one, as well as the composition property:

$$G_{mn} = G_m \circ G_n = G_n \circ G_m. \tag{8.4}$$

The 30–60–90 triangle is similar to the right isoceles triangle in being foldable not only into n^2 copies of itself but also into 3 (and therefore $3n^2$) copies of itself. Like the case of the right isoceles triangle, there also exist polynomials conjugate to these mappings. They are given by the definition:

$$G_{n\sqrt{3}}(h_G(u,v)) = h_G(nu - nv, nu + 2nv). \tag{8.5}$$

For example we have

$$G_{\sqrt{3}}(x,y) = (y, x^3 - 3xy - 9x - 6y - 12). \tag{8.6}$$

With definition (8.5), the composition formula (8.4) holds even if m or n or both are multiples of $\sqrt{3}$. Thus the recursion formulae (8.2) and (8.3) together with (8.6) allow us to calculate all of this "extra" sequence of polynomials.

The dynamical properties of the polynomials G_n can be obtained in the same way as those of the A_n; we list them here (n represents an integer or an integral multiple of $\sqrt{3}$):

(i) The function G_n has $\Pi(p, n)$ periodic orbits of period p in Δ_G, where Π is as in equation (6.1).

(ii) The Lyapunov exponents of G_n are both equal to $\log n$.

(iii) Almost all initial conditions in Δ_G have orbits which are dense in Δ_G, and the invariant measure generated by G_n is absolutely continuous with density

$$\delta(x, y) = \frac{3}{2\pi^2}\left[(4x^3 - y^2 - 12xy - 24y - 36x - 36)(4y + 12 - x^2)\right]^{-1/2}.$$

The polynomials G_n are also orthogonal with respect to this measure.

The fact that the equilateral triangle can be folded into two 30–60–90 triangles is reflected in the polynomials A_n and G_n. Define p: $\mathbf{C} \to \mathbf{R}^2$ by

$$p(z) = (z + \bar{z}, z\bar{z} - 3).$$

Then for each integer n, we have

$$p \circ A_n = G_n \circ p.$$

Acknowledgements. This research was supported by a grant from the Naval Academy Research Council.

I am indebted to Michael Hoffman for his invaluable contribution to the theory of folding polynomials.

REFERENCES

1. C. Berge, Principles of Combinatorics, Academic Press, New York, 1971.
2. R. Eier and R. Lidl, A Class of Orthogonal Variables in k Variables, *Math. Annalen*, 260 (1982) 93–99.
3. M. E. Hoffman and W. D. Withers, Generalized Chebyshev Polynomials Associated With Affine Weyl Groups (to appear).
4. T. Koornwinder, Two-Variable Analogues of the Classical Orthogonal Polynomials, in Theory and Application of Special Functions, ed. R. Askey, Academic Press, New York, 1975.
5. J. D. Lawrence, A Catalog of Special Plane Curves, Dover Publications, New York, 1972, pp. 131–135.
6. L. Mittag and M. J. Stephen, Yang-Lee Zeroes in the Potts Model, *J. Stat. Phys*, 35 (1984) 303–320.
7. T. J. Rivlin, The Chebyshev Polynomials, John Wiley & Sons, New York, 1974.

Celebrating Mathematics

LYNN ARTHUR STEEN, *St. Olaf College, Northfield, MN 55057*

LYNN STEEN is Past President of the Mathematical Association of America. This paper is the text of his Retiring Presidential Address presented January 8, 1988 at the annual meeting of the MAA in Atlanta. Steen received his B.A. from Luther College in 1961 and his Ph.D. from the Massachusetts Institute of Technology in 1965. He has studied at the Mittag-Leffler Institute and with his St. Olaf colleague J. Arthur Seebach, Jr. authored ***Counterexamples in Topology*** and served as co-editor of ***Mathematics Magazine***. Steen is editor of ***Mathematics Today, Mathematics Tomorrow***, and ***Calculus for a New Century: A Pump, Not a Filter***. He is Chairman of the Conference Board of the Mathematical Sciences and Chairman-elect of the Council of Scientific Society Presidents.

This is the year we celebrate American mathematics.

One hundred years ago, Thomas Scott Fiske founded the New York Mathematical Society, precursor to the American Mathematical Society. Then, mathematics flourished in Europe but barely existed in the New World. Today, the American Mathematical Society sustains the world's strongest environment for research mathematics. It has given science—and society—much to celebrate.

The Mathematical Association of America is nearing its 75th anniversary as an organization devoted to the teaching of mathematics, especially at the collegiate level. Ever since 1915, the Society and the Association have cooperated on many jointly sponsored activities. Our missions—mathematical research and mathematics teaching—are like braided strands that together form a strong fiber for the fabric of American science.

On behalf of the Mathematical Association of America, I salute our sister society for a century of accomplishment, for the creation of a rich tapestry of beautiful and useful mathematics. The AMS centenary beckons all of us in the mathematical community to look around to our colleagues, to look outward to society, and to look ahead to the future of mathematics. We join the celebration, to applaud the achievements of American mathematics; to proclaim that strong mathematics contributes to strong science, to strong defense, and to a strong economy; and to challenge American youth with the excitement of mathematical discovery.

Of course mathematics did not begin with the founding of the American Mathematical Society. Two hundred years before that—three hundred years ago—Newton published *Principia Mathematica*, thereby establishing mathematics as the methodological paradigm of theoretical science. From this paradigm have emerged many mathematical sciences, and many mathematics societies. The American Statistical Association celebrates its 150th

birthday in 1989; the Society for Industrial and Applied Mathematics celebrated its 35th Anniversary last year, at the same time that the Association for Computing Machinery celebrated its 40th birthday.

So "100 Years of American Mathematics" is a bit of hyperbole, a slogan that throws images beyond literal meaning. It is in part a strategy to focus attention on the need for sustained support for mathematics. We celebrate 100 Years of American Mathematics not because we equate Thomas Fiske with Isaac Newton, but because it is a timely device to set important issues before the mathematical world, the scientific community, and the attentive public. Despite a century of success, there are urgent matters that need attention.

Mathematics Today

Those of us who are part of the mathematical community recognize the enormous current vitality of the mathematical sciences. Applications, enthusiasm, initiatives, opportunities, and unity are the "vowels" around which mathematical language is formed, a language in which mathematicians express solutions to old problems and explore new areas of fruitful growth. Explosive growth is a sign of remarkable health, but it does leave the thousands of us who try to keep up panting breathlessly as the research leaders disappear over the horizon.

Dozens of areas of active research could be cited to document the vitality of contemporary mathematics (see [**17**], [**18**], [**24**]). For some of these, look at the program of this meeting. Better still, look at the whole program for 100 Years of American Mathematics, especially at the marvelous symposium that is part of the 1988 AAAS meeting in Boston and at the special series of expository lectures on contemporary mathematics that is featured at the AMS Centennial Meeting in Providence.

Today I want to highlight four areas as examples of the unity and applicability of mathematical research: computational statistics, mathematical biology, geometrical mathematics, and nonlinear dynamics. Each of these offers opportunities for initiatives in mathematical research and for engendering enthusiasm among students. Think especially about the latter, about opportunities for initiatives in mathematics education, while I briefly outline each of these four areas.

Computational Statistics. The statistical sciences study problems associated with uncertainty in the collection, analysis, and interpretation of data. Not surprisingly, the increasing use of computers to record and transform data has generated a host of new challenges for the statistical sciences.

For example, analysis of data from electronic scanning devices (in tomography, in aircraft or satellite reconnaissance, in environmental monitoring) has produced an urgent need for statistical analysis of data that has an

inherent spatial structure [**26**]. Research in this emerging field of spatial statistics employs a wide variety of mathematical, statistical, and computational techniques: problems of separating signals from noise borrow ideas from engineering; ill-posed scattering problems employ methods of numerical linear algebra; and smoothing of data requires statistical techniques of regularization. Underlying all this is the inherent geometry of the problem, which in many cases is dynamic and non-linear.

Many applications of statistics (for example, to clinical data from innovative medical protocols) involve small data sets from which one would like to infer meaningful patterns. Bradley Efron and others have pioneered innovative, computationally-intensive statistical techniques that use the limited available data to generate more data with the same statistical characteristics ([**4**], [**5**], [**14**]). By resampling the given data repeatedly, these so-called bootstrap methods generate millions of similar possible data sets which yield accurate approximations to various complex statistics. By comparing the value of statistics for the given sample with the distribution obtained by these resampling schemes, one can determine whether the observed values are significant.

Mathematical Biology. Nothing better illustrates the potential for mathematics in the biological sciences than the many traces of mathematics behind the Nobel prizes. For example, the 1979 Nobel Prize in medicine was awarded to Allan Cormack for his application of the Radon transform to the development of tomography and CAT scanners. The 1984 Nobel Prize in chemistry was awarded to biophysicist Herbert Hauptman, President of the Medical Foundation of Buffalo, for fundamental work in Fourier analysis pertaining to X-ray crystallography [**19**].

Indeed, recent research in the mathematical sciences suggests dramatically increased potential for fundamental advances in the life sciences using methods that depend heavily on mathematical and computer models. Structural biologists have become genetic engineers, capturing the geometry of complex macromolecules in supercomputers and then simulating interaction with other molecules in their search for biologically active agents. Using these computational methods, biologists can portray on a computer screen the geometry of a cold virus—an intricate polyhedral shape of uncommon beauty and fascinating geometric features—and search its surface for molecular footholds on which to secure their biological assault.

Geneticists are beginning the monumental effort to map the entire human genome, an enterprise requiring expertise in statistics, combinatorics, artificial intelligence, and data management to organize billions of bits of information. Ecologists—the first mathematical biologists—continue to use the extensive theories of population dynamics to predict the behavior and interaction of species ([**9**], [**32**]). Neurologists now use the theory of graphs to model networks of nerves in the body and the neural tangle in the brain [**10**].

Cell biologists study the replication of DNA using the newly discovered algebraic classification of knots ([**12**], [**13**]). Epidemiologists monitor the spread of AIDS with techniques that blend innovative statistics with classical analysis. And, finally, physiologists employ contemporary algorithms applied to nineteenth-century equations of fluid dynamics to determine such things as the effects of turbulence in the blood caused by cholesterol or swollen heart valves [**11**].

Geometrical Mathematics. Ever since Euclid, geometry has been one of the major pillars of core mathematics. After decades of decline (especially in mathematics teaching), the geometrical view in mathematics has undergone a renaissance, assisted both by the development of new theoretical tools and by the power of computer-based visual representation. In a very real sense, geometry is once again playing a central role on the stage of mathematics, much as it did in the Greek period.

Geometry claimed two of the three 1986 Fields Medals, which were awarded to Michael Freedman and Simon Donaldson for work in the geometry of four dimensional manifolds ([**1**], [**3**], [**21**]). By exploiting properties of the Yang-Mills field equations that reflect the wave-particle duality of matter, Donaldson showed that the differential geometry of four dimensional manifolds was vastly different than that suggested by their topological structure. Freedman provided the topological classification. Together, their work yielded not only deep understanding of four-dimensional manifolds, but the surprising insight that in four dimensions there are differentiable manifolds that are topologically but not differentiably equivalent to the standard Euclidean four dimensional space. Already, insight from this work has led to applications in string theory—the new super-symmetric theory of elementary particles—thereby providing fresh evidence (see [**34**]) of what Eugene Wigner called the "unreasonable effectiveness" of mathematics in the physical sciences.

Computer graphics provides a powerful new tool that extends geometrical techniques into many parts of mathematics. Computers—especially supercomputers—can calculate and display various mathematical structures in visual form, thereby enabling mathematicians to "see" the significance of abstract patterns that before could be interpreted only by formal means. For example, visual representations of solutions of differential equations often produce conjectures that open up whole new insights into the behavior of the system which the equations represent. Geometrical studies themselves regularly yield innovative challenges in the design of new algorithms and data structures, with spin-off benefits to applications in computer science (for example, database systems and word processing) far removed from the original geometric problem. The newly launched Geometry Supercomputing Project at the University of Minnesota is one example of the growing interaction of researchers in geometry with those in theoretical computer science.

Nonlinear Dynamics. Only in recent years have we been able to provide mathematical analyses of problems that are essentially nonlinear (for example, turbulence in fluids). These have been made possible by novel analytical methods, clever numerical simulation, and visual display on computer screens. Applications range from airfoil design to plasma physics, from oil recovery to studies of combustion ([**7**], [**15**], [**16**], [**22**]).

Nonlinear dynamics has yielded many surprises (see [**6**], [**8**], [**33**]), including long-term localized structures (e.g., the Red Spot on Jupiter), deterministic (rather than stochastically) generated chaotic motion (typical of some weather phenomena), and fractal patterns at the interface between fluids (for example, displacement of oil by water). The mathematics of nonlinear dynamics involves a great deal of traditional analysis (especially differential equations), reinforced by iterative processes, automata theory, and fractal geometry.

Computer display of nonlinear phenomena makes visible patterns that would never have been noticed by analytic means alone. In research on dynamical systems, on the transition from order to chaos, and on the emergence of fractal shapes from smooth flows, computers are to mathematics what telescopes and microscopes are to science: they increase by a thousandfold the portfolio of patterns that mathematicians can see and investigate.

The Newtonian revolution not only established mathematics as a paradigm for scientific reasoning, but it also established determinism as a paradigm for the behavior of physical systems. Nonlinear dynamics—a direct descendant of Newtonian mathematics—shows how ambiguity and uncertainty can arise in even simple deterministic systems, and how the onset of chaos itself can be predictable. In its power to change our Newtonian view of mathematics, nonlinear dynamics is as revolutionary as quantum mechanics: each breaks the bond of determinism and reveals entirely new structures that often defy what we have come to think of as common sense.

Mathematics in the Classroom

I choose these examples, drawn from diverse areas of the mathematical sciences, not just to illustrate the vitality of the mathematics that we are here to celebrate, but to provide a mathematical and intellectual backdrop for what is, unfortunately, a very different portrait of mathematics in the classroom.

In research we see a lot of geometry, a lot of data, a lot of science, a lot of computation—together with more traditional mathematical tools. We see investigation, exploration, and a continual search for pattern. Contemporary mathematics compels attention. It has the power to excite the best minds of our youth and to stimulate renewed creativity in teaching mathematics.

But this mathematics is *not* the mathematics taught in typical school or college classrooms. Far too often, mathematics in the classroom is a freeze-

dried mathematics—rigid, cold, and unappealing. Instead of exploration there is drill; instead of investigation, imitation. From elementary school arithmetic to college calculus, mathematics in the classroom is dramatically different from mathematics in practice.

You've all heard the litany of problems with school mathematics. It ranges from poor test performance on international assessments to declining interest among Americans in pursuing advanced study of mathematics. I won't repeat this evidence here since it has been widely publicized in reports, journals, and newsletters ([**20**], [**30**], [**31**]).

Not yet so well known are the current attempts by several organizations (for example, NCTM, MSEB, AAAS, and the University of Chicago) to reverse this decline ([**23**], [**27**], [**28**]). These projects have engaged school teachers, mathematics educators, and mathematics researchers in collaborative work on the problems of mathematics education in the schools. Although these projects differ greatly in purpose and detail, their emerging recommendations have much in common that resonates with the nature and practice of contemporary mathematics:

- Mathematics should be taught in a natural context;
- Students should be encouraged to create, to invent, and to participate;
- Calculators and computers should be used throughout the mathematics curriculum;
- New topics (for example, algorithms, data analysis, estimation) should be introduced into the mainstream curriculum;
- Facility in computation need not be a prerequisite to the study of mathematics;
- Mathematics should be studied as an integrated whole;
- Mathematics should help build students' abilities to reason logically;
- Communication is an important goal of mathematics instruction.

It doesn't take much imagination for someone who is familiar with examples of contemporary mathematical science to see how student involvement in such mathematics could contribute to achieving these goals. Just the examples I have cited—statistics, biology, geometry, dynamics—overflow with natural context and with opportunities for students to use computers to discover patterns. These examples reveal far better than the isolated morsels of the traditional curriculum that new mathematical methods are needed to solve new problems; that communication is important for one to just understand, let alone express, the subtleties revealed by mathematical analysis; and that mathematics in action requires not only calculation and logic, but also intuition, imagination, and organization.

Unfortunately, too few of those who are most knowledgeable about mathematical research are working with teachers to translate their research into experiences suitable for classroom exploration. And far too few teachers—even at the college and university level—have the background, the interest,

or the time to learn enough about modern mathematical research to translate it successfully into classroom experiences. So long as a large gap remains between those who create mathematics and those who teach mathematics, we cannot expect students to see in mathematics the challenge of an exciting and intellectually rewarding career.

Among the anniversaries we celebrate this winter is the centenary of the birth of George Pólya, who died just three years ago. Pólya was one of the rare mathematicians who made major contributions both to mathematics research and mathematics education. Andrei Kolmogorov, who died in October, was another.

The December 1987 issue of *Mathematics Magazine* is devoted to Pólya's life and work; I urge you to read it—I'm sure you'll find it as fascinating as I did. In that issue, Alan Schoenfeld wrote an interesting analysis of Pólya's theory of heuristics and its impact on teaching students to solve mathematical problems. Schoenfeld begins with Pólya's dictum that a good mathematics education is one that provides systematic opportunities for students to *discover* things.

How often does our teaching really do that? Think about the contrast of the stylized two-column proofs of high school geometry with the exploratory possibilities of three-dimensional computer graphics, or of the linking of knots, or of topological transformations of common surfaces. Or think, as many did at the NRC colloquium on Calculus for a New Century, about the contrast between the five thousand exercises in typical calculus books that mostly ask students to imitate calculator buttons, and the discovery potential in symbolic computer systems or in visual presentation of nonlinear dynamics.

Pólya's discovery dictum was echoed (perhaps unconsciously) at the calculus colloquium by Oberlin College President Frederick Starr [**29**]. He cited research [**2**] on college student career choices that shows "incontrovertibly" that the only institutions that are successfully resisting the precipitous decline in the percentage of students entering careers in science are those that base their pedagogy on a kind of apprenticeship system. In these schools students are brought into the laboratory to pursue real science under the direct gŭidance of professors who are themselves actively engaged in the scientific quest.

Traditionally, it has been the laboratory sciences—notably chemistry—that excel at attracting students by a style of education that involves students in the discovery of science. But now mathematics can do the same. With frontiers as exciting as chaotic systems and spatial statistics, there is no longer any reason for mathematics to fall behind the more glamorous laboratory sciences in attracting the interest and enthusiasm of our brightest youth.

Causes for Celebration

In celebrating mathematics, we point to the immense success of mathematical research in creating an intellectual understanding of space and number, of order and chaos, of pattern and disarray. Mathematics itself is beautiful, powerful, and deep; the process of doing mathematics is personally stimulating and intellectually rewarding.

Nevertheless, the profession of mathematics—as distinct from the discipline of mathematics—is not in good health. Decades of neglect in maintaining clear communication channels—with education, with science, with the public—have left mathematics isolated from the support systems that are vital to its health and well-being.

Our celebration must become a commitment to communicate. As we move into the second century of American mathematics—to continue the hyperbole—we should build on the impressive accomplishments in mathematics itself to bring the excitement and power of the mathematical sciences to all Americans. Here are five causes to champion as we celebrate mathematics in 1988.

- Involve Students in the Practice of Mathematics

The evidence is overwhelming that students receive a better education, and are more likely to be attracted to mathematics, when they are actively involved in mathematical experiences [**25**]. Pólya called it discovery learning; Starr described it as apprenticeship education. Although only a few mathematics students in the United States now receive the benefit of this type of learning, there are many excellent models of such teaching: problems competitions, research experiences for students, exploratory computer graphics, innovative tutorials where students become teachers, and team-based internships in mathematical modelling. It is noteworthy that the National Science Foundation, under the research directorates, has once again begun to support programs that provide research experiences to undergraduates. Especially for undergraduate students, but also in appropriate degrees for younger students, we must tilt the balance of mathematics education towards greater student involvement in learning.

Doing this will be expensive, and might involve radical departures in the way we finance undergraduate and graduate education. Since teachers tend to teach as they were taught, the most effective way to promote discovery learning in the schools is to enhance apprenticeship learning in the colleges, where tomorrow's teachers are today learning how to teach by the examples set by their college professors. The time is ripe for department chairs to insist that universities fund teaching at a sufficient level that all undergraduates can be taught by experienced teachers who will involve students in the excitement of discovering mathematics.

Apprenticeship education will require significantly better integration both

of research mathematics and of contemporary applications in the experience of students. Current research, new applications, and the emerging goals of mathematics education could resonate in ways that would greatly enhance student involvement in mathematical learning. But such resonance cannot happen so long as researchers and teachers continue to operate in separate spheres—worlds apart in mathematical outlook, experiences, and expectations. Resonance requires significant connection between stimulus and resonator. "Vertical integration" of mathematical knowledge that links research and applications with education, and that brings researchers into active contact with students and teachers, should become a major criterion in funding decisions that concern the support of research.

We should celebrate the wealth of interesting new mathematics by bringing this mathematics into every classroom in the nation. To do that will require changes in the way we judge teaching and in the way we judge research: each should be found wanting if it does not include appropriate linkage with the other. It is not enough that individuals be competent as teachers and separately as professionals: separate but equal is as inadequate as a model for the relation of teaching and professional activity as it is for racial composition of public schools. Our goal should be professional standards that insist on apprenticeship learning and vertical integration of mathematical research.

- EDUCATE THE ATTENTIVE PUBLIC

Far too many educated persons are ignorant of mathematics. More troublesome, most do not feel their mathematical ignorance to be a great handicap. Most successful lawyers, politicians, educators, business executives—and university administrators—have achieved positions of prominence with only a minimal (and frequently archaic) knowledge of mathematics. Moreover, many persons, whether well educated or not, harbor feelings of apprehension or even anxiety about mathematics due to an unpleasant early educational experience, often with something labelled the "new math." When we try to take the case for mathematics to the public—or even just to university administrators—we face not only the healthy skepticism that naturally greets any self-serving argument, but also ignorance, fear, and often hostility that is a legacy of our neglect of mathematics education.

The state of mathematics as a profession compels us to find ways to diminish public fear and ignorance of mathematics, for without broad public support—for teaching, for research, for encouragement of students—there is no possible way for the mathematical community on its own to sustain the momentum of the past half-century. Now, however, perhaps for the first time, the *breadth* of the mathematical landscape makes it possible at least to imagine overcoming this pervasive public apprehension of mathematics.

Mathematics now touches people's lives in ways that matter and that can be described and revealed in human terms. From symmetry and chaos

to computers and cosmology, from AIDS epidemiology and nuclear risks to political polls and ozone depletion, mathematics lurks behind most manifestations of science and technology. The same mathematization of society that makes the task of public understanding so essential also provides the means by which the task can be started—by building on mathematical ideas that are part of daily experience.

Surely part of our celebration must be to tell the story of mathematics to the attentive public. We are not publicists, but we are teachers. Suppose each department of mathematics made a commitment, just once each year, to arrange a public event that made mathematics visible in their community: an outside speaker who is working on something in which the public might be interested; a student project that involved a practical problem of interest to the community; a forum on the changing nature of school mathematics; or an exposition of a slice of mathematics related to some professor's research. Since the public is always more interested in people than in abstractions, there are, in addition, many good opportunities for news stories in hometown papers about the accomplishments of students. Someday some mathematics department should try to put out as many publicity releases on the accomplishments of their students as the athletic department does of theirs.

▪ EXPLORE FUNDAMENTAL ISSUES IN MATHEMATICS EDUCATION

Computers influence mathematicians not only by providing new tools for research and teaching, but also by posing deep questions about central issues in our discipline. Now that calculators can manipulate symbols and calculate answers,

- What—if not arithmetic—should be the core of elementary school mathematics?
- What—if not manipulation—should be the core of high school algebra?
- What—if not calculation—should be the core of calculus?
- What—if not calculus—should be the core of college mathematics?

At the same time that computers force attention on issues that are deeply rooted in unexamined tradition, mathematical research has transformed the nature of mathematics, opening up new options for what might be considered central and what derivative among the concepts of mathematics.

We need to find new threads of continuity with which to weave a mathematics curriculum for the twenty-first century. Finding appropriate central themes poses an immense challenge for the best minds among us, researchers and teachers alike. It gives common purpose to our diverse expertise, and sets a common agenda for those in research, those in college teaching, and those in school mathematics.

This provides yet another occasion for celebration: the opportunity, joined with the need, to transform school mathematics in ways that reflect the richness and diversity of mathematical research and mathematical applications.

Vertical integration of research, applications, and teaching will help bring about this transformation. But we need, in addition, structured opportunities for reflection in which the most synoptic thinkers among us bring their experiences in research, in applications, and in teaching to bear on the task of articulating central themes for mathematics education as we move into the next century.

- Ensure for All Students
  Equal Opportunities for Mathematical Success

Despite the culturally neutral status of mathematics (as compared, say, to biology), the last decade has produced distressing evidence of class and ethnic distinctions arising as a result of the way mathematics education is practiced in the United States. One-third of U.S. students—Blacks, Hispanics, and Native Americans—provides fewer than 10% of mathematics graduates, despite the evidence from isolated model programs that excellent retention and success rates can be achieved within a suitable educational context. Another third of U.S. students—white females—drops out of advanced degree programs in the mathematical sciences at twice the rate of male students.

Compounding these problems of class distinctions are the political, educational, and social side effects of large numbers of foreign-born teaching assistants in our major universities. It is easy to make a strong case for having many foreign graduate students in our universities; I join the many scientific leaders who defend this practice which has led to the United States being, in the words of Robert White, the "schoolhouse of the world." It is harder to make a case for placing inexperienced foreign graduate students in the classroom as instructors for American students who are not prepared to cope simultaneously with the challenge of a foreign culture, a foreign language, and a foreign discipline—namely, mathematics.

The consequence of these two unrelated trends is that both majority and minority students in college classes often receive mathematics instruction in a context that is culturally alien to them. Students who have too little in common with their teachers are unable to see themselves as future mathematicians or mathematics teachers. In this context, it is not surprising that even white U.S. males are no longer choosing careers in mathematical sciences. As I am sure you are well aware, the number of U.S. males receiving Ph.D. degrees each year in mathematics is less than 40% of what it was fifteen years ago.

The problem of opportunity in mathematics is so serious and so difficult that it is hard to even imagine a solution that is feasible, much less optimal. However, the mathematical community has an enormous resource that can be brought to bear on this problem—namely, the strong and culturally diverse community of research mathematicians that proves by its very existence the universality of mathematics. We need to find effective ways

of conveying the rich and worldwide nature of mathematics to youth from the many subcultures that contribute to the American mosaic, and then to provide a context appropriate to their backgrounds in which to nurture mathematical talent.

The Professional Development Program at at Berkeley, led by Uri Treisman and Leon Henkin, has transformed the success rates of minority freshmen and enabled many to finish Berkeley and proceed to advanced or professional degrees. The University of Michigan has had considerable success in interesting minority students in careers in science through a special program that provides experiences in undergraduate research. These examples—and I'm sure there are others—show that it *is* possible to successfully attract talented minority students to careers in mathematics and science. Making progress in this endeavor would provide just cause for a true celebration.

- INVEST IN TODAY'S EDUCATION TO STRENGTHEN TOMORROW'S RESEARCH

Current debate about support for mathematics too often pits research against education, when in reality today's education is the pipeline for tomorrow's research. We read in the *Notices* of mathematicians who are under pressure to get research grants on pain of "being fired, having their teaching loads raised, or not getting raises." We hear continuing concern in the research community that, in times of limited budgets, new funds for education might be subtracted from the already limited amounts available for support of basic research—despite the fact that the percentage of federal support for science and mathematics that goes to education has slipped during the past four decades from nearly 50% to around 10%.

These arguments follow the same pattern that has led to the enormous growth in the federal deficit: by giving priority to the immediate needs of those in positions of power, we in effect support adults at the expense of children. Mathematicians, of all people, should be able to plan strategies that will optimize the strength of mathematical and scientific research over the long term. Part of that strategy is the recognition that education is not an alternative to research, but the foundation for future research.

Our celebration of the bounty of mathematical research must entail a commitment to education as the wellspring of research. We need to use multidimensional criteria in deciding on priorities for our community, seeking strategies that lead simultaneously to improvement in school mathematics, in collegiate mathematics, in graduate education, and in research.

All One System

Despite appearances to the contrary, mathematical research is inextricably entwined with mathematics education at all levels, with science and engineering, and with political, economic, and sociological aspects of society

at large. Educators and researchers, teachers and professors, mathematicians and scientists—we are all part of a single system of knowledge on which contemporary society depends.

It is in this spirit that we join in celebrating the centenary of the American Mathematical Society. "100 Years of American Mathematics" provides the occasion for building new mathematical science on the firm foundation of the past century's research. Our centenary causes should be as broad and sweeping as our discipline:

- To involve students in the practice of mathematics.
- To educate the attentive public.
- To explore fundamental issues in mathematics education.
- To ensure for all students equal opportunity for mathematical success.
- To invest in today's education to strengthen tomorrow's research.

What transforms these causes from empty rhetoric to concrete options is the opportunity for education and communication implicit in the advances of today's mathematical sciences—in such areas as computational statistics, mathematical biology, geometrical mathematics, and nonlinear dynamics. It is in the frontiers of mathematical science—not in current textbooks or today's classrooms—that one can find the innovative and intellectually rewarding options needed to transform education, to excite our youth, to educate the public, and to reach all Americans.

This should be our centennial cause, not just for 1988 but for the rest of this century. It is a cause that can unite researchers and educators in a common challenge: To let the power and beauty of mathematics speak for itself.

REFERENCES

1. Michael Atiyah, On the Work of Simon Donaldson, *Proceedings of the International Congress of Mathematicians, 1986,* American Mathematical Society, 1988, pp. 3-6.
2. David Davis-Van Atta, et al., *Educating America's Scientists: The Role of the Research College,* Oberlin College, 1985.
3. Simon K. Donaldson, The Geometry of 4-Manifolds, *Proceedings of the International Congress of Mathematicians, 1986,* American Mathematical Society, 1988, pp. 43-54.
4. Bradley Efron, *The Jackknife, the Bootstrap and Other Resampling Plans,* Society for Industrial and Applied Mathematics, 1982.
5. ———, Bootstrap and Other Resampling Methods, in *Mathematical Sciences: Some Research Trends,* National Academy of Sciences, 1988.
6. Mitchell J. Feigenbaum and Martin Kruskal, Order, Chaos, and Patterns: Aspects of Nonlinearity, *Research Briefings 1987,* National Academy of Sciences, 1988.
7. James Gleik, *Chaos,* Viking Press, 1987.
8. Celso Grebogi, Edward Ott, and James A. Yorke, Chaos, Strange Attractors, and Fractal Basin Boundaries in Nonlinear Dynamics, *Science,* 238 (30 October 1987) 632-638.
9. Thomas C. Hallam and Simon Levin, Editors, *Mathematical Ecology: An Introduction,* Springer-Verlag, 1986.
10. Frank C. Hoppensteadt, *An Introduction to the Mathematics of Neurons,* Cambridge University Press, 1986.

11. ———, Editor, *Mathematical Aspects of Physiology,* Lectures in Applied Mathematics, V. 19, American Mathematical Society, 1981.
12. Vaughan F.R. Jones, A New Knot Polynomial and von Neumann Algebra, *Notices of the Amer. Math. Soc.,* 33 (March 1986) 219-225.
13. Gina Kolata, Solving Knotty Problems in Math and Biology, *Science,* 231 (28 March 1986) 1506-1508.
14. ———, The Art of Learning from Experience, *Science,* 225 (13 July 1984) 156-158.
15. Mort La Brecque, Fractal Symmetry, *Mosaic,* 16:1 (Spring 1985).
16. ———, Fractal Applications, *Mosaic,* 17:4 (Winter 1986/7); 18:1 (Spring 1987).
17. *Mathematical Sciences: A Unifying and Dynamic Resource,* National Academy of Sciences, 1986.
18. *Mathematical Sciences: Some Research Trends,* National Academy of Sciences, 1988.
19. *Mathematics: The Unifying Thread in Science,* Notices of the Amer. Math. Soc., 33 (1986) 716-733.
20. Curtis C. McKnight, et al., *The Underachieving Curriculum: Assessing U.S. School Mathematics from an International Perspective,* Stipes Publishing Company, 1987.
21. John Milnor, The Work of Michael Freedman, *Proceedings of the Interantional Congress of Mathematicians, 1986,* American Mathematical Society, 1988, pp. 13-15.
22. Ivars Peterson, Packing It In: Fractals Play an Important Role in Image Compression, *Science News,* 131 (2 May 1987) 283-285.
23. Anthony Ralston, et al., *A Framework for Revision of the K-12 Mathematics Curriculum,* Task Force Report submitted to the Mathematical Sciences Education Board, National Research Council, January 1988.
24. *Renewing U.S. Mathematics: Critical Resource for the Future,* National Academy of Sciences, 1984.
25. Lauren B. Resnick, *Education and Learning to Think,* National Academy Press, 1987.
26. Werner C. Rheinboldt, *Future Directions in Computational Mathematics, Algorithms, and Scientific Software,* Society for Industrial and Applied Mathematics, 1985.
27. Thomas Romberg, et al., *Curriculum and Evaluation Standards for School Mathematics,* Working Draft, National Council of Teachers of Mathematics, October 1987.
28. James Rutherford, et al., *What Science is Most Worth Knowing?* Draft Report of Phase I, Project 2061; American Association for the Advancement of Science, December 1987.
29. Lynn Arthur Steen, Editor, *Calculus for a New Century: A Pump, Not a Filter,* Mathematical Association of America, 1988.
30. ———, Mathematics Education: A Predictor of Scientific Competitiveness, *Science,* 237 (17 July 1987) 251-252, 302.
31. Harold W. Stevenson, et al., Mathematics Achievement of Chinese, Japanese, and American Children, *Science,* 231 (14 February 1986) 693-699.
32. E. Teramoto and M. Yamaguti, *Mathematical Topics in Population Biology, Morphogenesis, and Neurosciences,* Lecture Notes in Biomathematics, Vol. 71, Springer-Verlag, 1987.
33. M. Mitchell Waldrop, Computers Amplify Black Monday, *Science,* 238 (30 October 1987) 602-604.
34. Edward Witten, Physics and Geometry, *Proceedings of the International Congress of Mathematicians, 1986,* American Mathematical Society, 1988, pp. 267-303.

Unsolved Problems

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

Has Every Latin Square of Order n a Partial Latin Transversal of Size $n-1$?

P. ERDŐS
Hungarian Academy of Sciences, Realtanoda u. 13-15, Budapest V, Hungary

D. R. HICKERSON, D. A. NORTON, AND S. K. STEIN
Department of Mathematics, University of California at Davis, Davis, CA 95616

The notion of a transversal of a latin square easily generalizes to more general arrays. We present some of the known combinatorial results in this area and indicate the open questions.

They concern $m \times n$ rectangular arrays of mn cells ($m \le n$), each cell containing a symbol. A **transversal** of an array is a set of m cells, no two in the same row or same column. A **latin transversal** is one whose symbols are distinct. A row or column is **latin** if its symbols are distinct. If $m = n$ and each symbol occurs exactly n times, we call the array an **equi-n-square**. If each row and column of an equi-n-square is latin, it is a **latin square**. The definitive work [**1**] on the subject, published in 1974, contains 639 references and 73 problems. The earliest result is due to Euler.

Marshall Hall [**2**, and see the earlier work of Lowell Paige, **3**] examined transversals of $n-1$ by n arrays obtained from an abelian group A of order n, as follows. Let $A = \{a_1, \ldots, a_n\}$. Corresponding to each element a in A is the sequence $\{a + a_1, \ldots, a + a_n\}$, recording translation by a. Let $\{b_1, \ldots, b_{n-1}\}$ be a sequence of $n-1$ not necessarily distinct elements of A, and form the $n-1$ by n array of the translations of A by $b_1, \ldots, b_{n-1}$. Each row is latin, but there may be duplications in the columns. Hall proved that such an array has a latin transversal.

In [**5**] Stein introduced several types of arrays, including the equi-n-square, and established the existence of transversals with many distinct elements. For example, the equi-n-square has a transversal with at least

$$n\left(1 - \frac{1}{2!} + \frac{1}{3!} - \cdots \pm \frac{1}{n!}\right) \approx (1 - 1/e)n \approx 0.63n$$

distinct symbols. He also gave equi-n-squares ($n \ge 2$) without a latin transversal, and showed that n by n arrays in which each symbol appears exactly q times have transversals with at least $n - q/2$ symbols.

More recently, Shor [4] showed that every latin square has a partial latin transversal of length at least

$$n - 5.53(\ln n)^2$$

and observed that his method will yield better results, but not as good as $n - \log_2 n$ (log to base 2).

There are several outstanding conjectures. The most noteworthy are:

Conjecture 1. *An equi-n-square has a transversal with at least $n-1$ distinct symbols.*

The special case of the title has been associated with Herb Ryser's name, while [**1**, p. 103] attributes it to Richard Brualdi. The answer is affirmative for latin squares which are the Cayley tables of abelian groups.

Conjecture 2. *An $n-1$ by n array in which each symbol appears at most q times ($q \leq n$) has a latin transversal.*

The case $q = n$ would imply that a row-latin $n-1$ by n array has a latin transversal, and would also imply Conjecture 1. For $q = 1$, conjecture 2 is trivally true, and the case $q = 2$ follows from $q = 3$, which can be proved as follows.

THEOREM 1. *Let $2 \leq k < m \leq n$ and assume that every m by n array in which each symbol appears at most 3 times has a latin transversal that misses k proscribed cells. Then every such $m+1$ by $n+1$ array has a latin transversal that misses k proscribed cells.*

Proof. Consider an $m+1$ by $n+1$ array in which each symbol appears at most 3 times and in which k cells are proscribed. There are two cases: (1) some row or column contains at least 2 proscribed cells; (2) no row or column contains more than 1 proscribed cell.

Case 1. Consider a row (or column) with at least 2 proscribed cells. Since $k < m$, there is a cell c in that row which is not proscribed. Denote the symbol in that cell by t. Delete the row and column containing c, producing an m by n array. This array has at most $k-2$ proscribed cells and at most 2 cells with the symbol t, so it has a latin transversal that misses the proscribed cells and does not have the symbol t in it. Adjunction of cell c gives a latin transversal for the original $m+1$ by $n+1$ array.

Case 2. Consider a cell that is on a row of one proscribed cell and on a column of another proscribed cell. Delete the row and column of this cell and argue as in Case 1.

An exhaustive search showed that every 4 by 5 array in which each symbol occurs at most 3 times has a latin transversal avoiding any 2 proscribed elements. Hence every m by $m+1$ array has this property for $m \geq 4$. A similar analysis showed that for $n \geq 6$, every n by n array in which each symbol appears at most 3 times possesses a latin transversal avoiding 2 proscribed cells.

Theorem 2 is proved much as Theorem 1.

THEOREM 2. *If every m by n array in which each symbol occurs at most 3 times has a latin transversal that passes through any given cell, then every $m+1$ by $n+1$ array in which each symbol appears at most 3 times has the same property.*

Note that there is no latin transversal through the top left cell of a 4 by 4 array which contains the configuration of FIG. 1.

Our final theorem is an unpublished result of Erdős and Joel Spencer.

a			
	a	b	b
	b	a	c
	c		

FIG. 1.

THEOREM 3. *Let k be fixed. An $n \times n$ array in which each symbol appears at most k times has a latin transversal if n is sufficiently large.*

Their proof will also work if $k < (\ln n)^{1-\varepsilon}$ and they believe that it may be possible to make it do so for $k < (\ln n)^c$ for every fixed c, but beyond this there may be serious difficulties. [J. Dénes **[6]** made a conjecture equivalent to Conjecture 1, and some other relevant remarks.—Ed.]

REFERENCES

1. J. Dénes and A. D. Keedwell, Latin Squares and their Applications, Akadémiai Kiadó, Budapest & English Universities Press, 1974.
2. M. Hall, A combinatorial problem on abelian groups, *Proc. Amer. Math. Soc.*, 3 (1952) 584–587; MR 14, 350.
3. L. J. Paige, A note on finite abelian groups, *Bull. Amer. Math. Soc.*, 53 (1947) 590–593; MR 9, 6c.
4. P. W. Shor, A lower bound for the length of a partial transversal in a latin square, *J. Combin. Theory Ser.* A., 33 (1982) 1–8; MR 83j:05017.
5. S. Stein, Transversals of latin squares and their generalizations, *Pacific J. Math.*, 59 (1975) 567–575; MR 52 #7930.
6. J. Dénes, Research problem 40, *Period. Math. Hungar.*, 17 (3) (1986) 245–246.

NOTES

EDITED BY DENNIS DETURCK, RICHARD LIBERA, AND ANITA E. SOLOW

More on Fresnel Integrals

I. E. LEONARD
Department of Mathematical Sciences, Northern Illinois University, DeKalb, IL 60115

In [2], Flanders evaluates the Fresnel integrals

$$F_0 = \int_0^\infty \cos x^2\, dx \quad \text{and} \quad G_0 = \int_0^\infty \sin x^2\, dx$$

using a strictly real-variable derivation. He considers, for $t \geqq 0$,

$$F(t) = \int_0^\infty e^{-tx^2} \cos x^2\, dx \quad \text{and} \quad G(t) = \int_0^\infty e^{-tx^2} \sin x^2\, dx$$

and shows that they satisfy the functional equations

$$F(t)^2 - G(t)^2 = \frac{1}{4}\pi \frac{1}{1+t^2} \quad \text{and} \quad 2F(t)G(t) = \frac{1}{4}\pi \frac{1}{1+t^2},$$

which he then solves to obtain

$$F(t) = \sqrt{\frac{\pi}{8}} \sqrt{\frac{\sqrt{1+t^2}+t}{1+t^2}}, \qquad G(t) = \sqrt{\frac{\pi}{8}} \sqrt{\frac{\sqrt{1+t^2}-t}{1+t^2}}.$$

Towards the end of the article, Flanders remarks that replacing x^2 by x yields the Laplace transforms of $(1/2)(\cos x/\sqrt{x})$ and $(1/2)(\sin x/\sqrt{x})$, respectively, that is,

$$F(t) = \frac{1}{2}\int_0^\infty e^{-tx} \frac{\cos x}{\sqrt{x}}\, dx \quad \text{and} \quad G(t) = \frac{1}{2}\int_0^\infty e^{-tx} \frac{\sin x}{\sqrt{x}}\, dx. \tag{1}$$

In this note, we show that F and G can be transformed into improper integrals which may be evaluated directly, using only the method of partial fractions; see, e.g. [3].

For $t > 0$,

$$\begin{aligned} G(t) &= \frac{1}{2}\int_0^\infty \frac{\sin x}{\sqrt{x}} e^{-xt}\, dx = \frac{1}{2}\int_0^\infty \sin x\, e^{-xt}\left(\frac{1}{\sqrt{\pi}}\int_0^\infty \frac{1}{\sqrt{s}} e^{-sx}\, ds\right) dx \\ &= \frac{1}{2\sqrt{\pi}}\int_0^\infty \left(\int_0^\infty \frac{\sin x}{\sqrt{s}} e^{-(s+t)x}\, ds\right) dx \\ &= \frac{1}{2\sqrt{\pi}}\int_0^\infty \left(\frac{1}{\sqrt{s}}\int_0^\infty \sin x\, e^{-(s+t)x}\, dx\right) ds \\ &= \frac{1}{2\sqrt{\pi}}\int_0^\infty \frac{ds}{\sqrt{s}\left(1+(s+t)^2\right)} = \frac{1}{\sqrt{\pi}}\int_0^\infty \frac{du}{1+(u^2+t)^2}. \end{aligned}$$

Therefore,

$$G(t) = \frac{1}{\sqrt{\pi}}\int_0^\infty \frac{du}{1+(u^2+t)^2} \tag{2}$$

and similarly,

$$F(t) = \frac{1}{\sqrt{\pi}} \int_0^\infty \frac{(u^2 + t)\, du}{1 + (u^2 + t)^2} \tag{3}$$

for $t > 0$.

The interchange of the order of integration is easily justified, since $f(s, x) = (\sin x/\sqrt{s})e^{-(s+t)x}$ is continuous for $0 < a \leqq s \leqq b$, $0 \leqq x < \infty$, and $\int_0^\infty f(s, x)\, dx$ converges uniformly for s in $[a, b]$. (see, e.g. **[1]**, p. 207). Use was also made of the elementary Laplace transforms

$$\int_0^\infty \frac{1}{\sqrt{s}} e^{-sx}\, ds = \sqrt{\frac{\pi}{x}}, \qquad \int_0^\infty \sin x\, e^{-(s+t)x}\, dx = \frac{1}{1 + (s+t)^2},$$

and

$$\int_0^\infty \cos x\, e^{-(s+t)x}\, dx = \frac{s+t}{1 + (s+t)^2}.$$

Now we factor $1 + (u^2 + t)^2$ over $\mathbb{R}$ into (irreducible) factors:

$$1 + (u^2 + t)^2 = (u^2 - 2au + b)(u^2 + 2au + b),$$

where

$$a = \frac{1}{\sqrt{2}}\left[\sqrt{1 + t^2} - t\right]^{1/2} \quad \text{and} \quad b = \sqrt{1 + t^2}.$$

The partial fraction decompositions of the integrands in (2) and (3) are easily found to be

$$\begin{aligned} \frac{1}{1 + (u^2 + t)^2} = &-\frac{1}{8ab}\frac{2(u-a)}{u^2 - 2au + b} + \frac{1}{4b}\frac{1}{(u-a)^2 + b - a^2} \\ &+ \frac{1}{8ab}\frac{2(u+a)}{u^2 + 2au + b} + \frac{1}{4b}\frac{1}{(u+a)^2 + b - a^2} \end{aligned}$$

and

$$\begin{aligned} \frac{u^2 + t}{1 + (u^2 + t)^2} = &\frac{b-t}{8ab}\frac{2(u-a)}{u^2 - 2au + b} + \frac{b+t}{4b}\frac{1}{(u-a)^2 + b - a^2} \\ &- \frac{b-t}{8ab}\frac{2(u+a)}{u^2 + 2au + b} + \frac{b+t}{4b}\frac{1}{(u+a)^2 + b - a^2}. \end{aligned}$$

Integrating, we obtain

$$\begin{aligned} \int \frac{du}{1 + (u^2 + t)^2} = &\frac{1}{8ab}\log\left(\frac{u^2 + 2au + b}{u^2 - 2au + b}\right) \\ &+ \frac{1}{4b\sqrt{b - a^2}}\left[\tan^{-1}\left(\frac{u-a}{\sqrt{b - a^2}}\right) + \tan^{-1}\left(\frac{u+a}{\sqrt{b - a^2}}\right)\right], \end{aligned}$$

and

$$\int \frac{(u^2+t)\,du}{1+(u^2+t)^2} = \frac{b-t}{8ab}\log\left(\frac{u^2-2au+b}{u^2+2au+b}\right)$$
$$+\frac{b+t}{4b\sqrt{b-a^2}}\left[\tan^{-1}\left(\frac{u-a}{\sqrt{b-a^2}}\right)+\tan^{-1}\left(\frac{u+a}{\sqrt{b-a^2}}\right)\right].$$

We can now evaluate the integrals (2) and (3) and obtain (1) for $t > 0$. Since F and G are continuous on $[0, \infty)$, letting $t \to 0^+$, we have

$$F_0 = \int_0^\infty \cos x^2\,dx = \sqrt{\frac{\pi}{8}} \quad \text{and} \quad G_0 = \int_0^\infty \sin x^2\,dx = \sqrt{\frac{\pi}{8}}$$

as required.

REFERENCES

1. R. C. Buck, Advanced Calcúlus, 2nd Edition, McGraw-Hill, New York, 1965, p. 207.
2. H. Flanders, On the Fresnel integrals, The American Mathematical Monthly, 89 (1982) 264–266.
3. D. V. Widder, Advanced Calculus, 2nd Edition, Prentice-Hall Inc., Englewood Cliffs, N.J., 1961, p. 382.

A Class of Abelian Groups Arising from an Analysis of a Proof

GARY A. MARTIN
Mathematics Department, Purdue University, Calumet Campus, Hammond, IN 46323

The 1973 Putnam Competition asks for a proof of the fact that if $2n + 1$ integers have the property that no matter which integer is deleted, some n of those remaining have the same sum as the other n, then all of the original $2n + 1$ numbers are equal. Even before beginning to try to prove this, I was struck by the hypothesis that the numbers are integers, for the theorem seems equally plausible for rational, real, and complex numbers. Indeed, the assertion is *meaningful* for any abelian group. As I will demonstrate below, it is *true* for precisely those abelian groups which have no nontrivial elements of odd order. The proof is a generalization and extension of the slick proof of the original statement, given in [**1**].

While trying to solve the problem for integers, I made two useful observations, both of which provided evidence that the assertion was indeed true in greater generality. First, one can generalize immediately to rationals by multiplying each number by a common multiple of the denominators, applying the known result for integers, and then dividing. This idea, that the truth of the hypothesis and the truth of the conclusion are invariant under a basic operation, reappears in Lemma 7 and is one of two ideas that constitute the proof given in [**1**]. Second, if $2n + 1$ complex numbers satisfy the hypothesis of the problem, then so do their real and imaginary parts. If one assumes that the statement is true of real numbers, then the $2n + 1$ real parts are equal and the $2n + 1$ imaginary parts are equal, so the complex numbers are all equal. In a more general setting, this is just the observation that the class of abelian groups for which the statement is true is closed under direct products. This

suggests that one should consider the problem from a universal-algebraic point of view.

We will use the following notation: For each n and any $x_1, \ldots, x_{2n+1}$ in an abelian group, let $\varphi_n(x_1, \ldots, x_{2n+1})$ be the statement of the hypothesis, namely, that for each k, there is a partition of the indices excluding k into two classes each of size n for which the sum of the x_j for j in one class equals the sum of the x_j for j in the other class. For each n, let $\psi_n(x_1, \ldots, x_{2n+1})$ be the statement that $x_1 = x_2 = \cdots = x_{2n+1}$. For each n, let θ_n be the statement

$$\forall x_1, \ldots, x_{2n+1}(\varphi_n(x_1, \ldots, x_{2n+1}) \Rightarrow \psi_n(x_1, \ldots, x_{2n+1})).$$

Then $\varphi_n(x_1, \ldots, x_{2n+1})$ and $\psi_n(x_1, \ldots, x_{2n+1})$ are statements about $2n + 1$ elements of an abelian group, whose truth depends on both the choice of those elements and the group. θ_n is a statement about abelian groups; if θ_n is true in the abelian group A, then one says that A *satisfies* θ_n.

We begin by examining the closure properties of the class of abelian groups which satisfy any particular θ_n.

LEMMA 1. *If A is an abelian group satisfying θ_n, and if B is a subgroup of A, then B satisfies θ_n.*

The proof is an easy exercise for the reader.

LEMMA 2. *If A is an abelian group, and every finitely generated subgroup of A satisfies θ_n, then A satisfies θ_n.*

Proof. Fix $x_1, \ldots, x_{2n+1}$ in A. If $\varphi_n(x_1, \ldots, x_{2n+1})$ is true in A, then it is also true in the subgroup of A generated by $x_1, \ldots, x_{2n+1}$. Then $\psi_n(x_1, \ldots, x_{2n+1})$ is true in that subgroup, and hence in A as well.

COROLLARY 3. *An abelian group satisfies θ_n if and only if each of its finitely generated subgroups does.*

LEMMA 4. *If θ_n is satisfied by each of the abelian groups $A_1, \ldots, A_n$, then θ_n is satisfied by the direct sum $A = \oplus A_k$.*

Proof. Given $x_1, \ldots, x_{2n+1}$ in A, let $x_{j,k}$ denote the kth coordinate of x_j. Suppose $\varphi_n(x_1, \ldots, x_{2n+1})$ is true in A. Then $\varphi_n(x_{1,k}, \ldots, x_{2n+1,k})$ is true in A_k, since the partition of the indices which works for A will also work for A_k. Then by hypothesis, $\psi_n(x_{1,k}, \ldots, x_{2n+1,k})$ is true in A_k for each k, so $\psi_n(x_1, \ldots, x_{2n+1})$ is true in A.

COROLLARY 5. *An abelian group A satisfies θ_n if and only if each cyclic subgroup of A satisfies θ_n.*

Proof. Suppose that each cyclic subgroup of A satisfies θ_n. Any finitely generated subgroup B of A is a direct sum of cyclic subgroups of A by the fundamental theorem of finitely generated abelian groups. By Lemma 4, B satisfies θ_n. By Corollary 3, A satisfies θ_n. The converse follows from Lemma 1.

It suffices, then, to consider cyclic groups. The style of the statements changes as the central issue comes more clearly into focus.

LEMMA 6. *If a cyclic group A satisfies every θ_n, then either A is infinite or A has order 2^m for some m.*

Proof. If the conclusion fails, then A has an element a of order n, where $n > 1$ is odd. Let $x_1 = \cdots = x_n = 0$ and $x_{n+1} = \cdots = x_{2n+1} = a$. Then if a zero is deleted, half of the $n + 1$ a's may be paired with half of the $n - 1$ zeros to obtain equal sums, and if an a is deleted, the remaining n a's add to zero, as do the n zeros. This demonstrates the failure of θ_n in A.

LEMMA 7. *If $\varphi_n(x_1, \ldots, x_{2n+1})$ holds for $x_1, \ldots, x_{2n+1}$ in an abelian group A, then for any $a \in A$, $\varphi_n(x_1 + a, \ldots, x_{2n+1} + a)$ holds in A as well.*

Proof. For each k, the partition of the x_j's which works when x_k is deleted also works for the $x_j + a$'s when $x_k + a$ is deleted.

LEMMA 8. *If B is a subgroup of index 2 in an abelian group A, and if $\varphi_n(x_1, \ldots, x_{2n+1})$ holds for $x_1, \ldots, x_{2n+1} \in A$, then $x_1 \equiv \cdots \equiv x_{2n+1} (\mathrm{mod}\ B)$.*

Proof. For any $x \in A$, $2x \in B$. For any k, the sum of all of the x_j's is x_k plus twice the sum of some n of the other x_j's. Thus each x_k is congruent modulo B to the sum of all of the x_j's, and hence all of the x_k's are congruent to each other.

LEMMA 9. *If A is infinite cyclic or cyclic of order 2^m, then A satisfies every θ_n.*

Proof. In either case, A has a descending chain of subgroups, $A \geqslant 2A \geqslant 4A \geqslant \cdots$, each of index 2 in the preceding, whose intersection is the trivial subgroup. Let $x_1, \ldots, x_{2n+1} \in A$, and suppose $\varphi_n(x_1, \ldots, x_{2n+1})$ holds. Then by Lemma 7, $\varphi_n(0, x_2 - x_1, \ldots, x_{2n+1} - x_1)$ holds as well. By Lemma 8, each of $x_2 - x_1, \ldots, x_{2n+1} - x_1$ lies in each of the subgroups in the descending chain. Thus $x_1 = x_2, \ldots, x_1 = x_{2n+1}$.

(The proof of Lemma 9 for infinite cyclic groups is essentially that appearing in [**1**].)

The classification of abelian groups which satisfy every θ_n is now complete.

THEOREM 10. *An abelian group satisfies every θ_n if and only if it has no nontrivial elements of odd order.*

Proof. This follows immediately from Corollary 5, Lemma 6, and Lemma 9.

There is another natural direction in which to generalize the problem. Fix an integer $r > 1$. Consider the class C_r of all abelian groups A for which for every n the following statement is true: If $rn + 1$ elements of A have the property that no matter which one is deleted, the rest can be partitioned into r sets of size n all with the same sum, then all $rn + 1$ elements must be equal. Again, the problem is to find a description of this class of abelian groups.

In direct analogy to Lemma 6, if A has an element a whose order $n > 1$ is relatively prime to r, then one sees that A is not in the class C_r as follows. Let $0 < m < r$ be a solution to $mn \equiv 1 \pmod r$. Take mn of the elements to be 0 and the remaining $(r - m)n + 1$ to be a. Then if a 0 is deleted, both the number of remaining 0's and the number of remaining a's are divisible by r, so one may take the same number of 0's and a's in each block of the partition. If an a is deleted, then take m sets of n zeros and $r - m$ sets of n a's.

Lemmas 1, 2, and 4, and Corollaries 3 and 5 generalize easily to this setting, with the consequence that an abelian group is in C_r if and only if all of its cyclic subgroups are. By the example in the preceding paragraph, it suffices to analyze the

infinite cyclic group and cyclic groups of order p^m, where p is a prime which divides r. It is easy to see that Lemmas 7 and 8 remain true when 2 is replaced by r and the subgroup B is taken to have index p in A. Then Lemma 9 generalizes since both the infinite cyclic group and the cyclic groups of order p^m have decreasing sequences of subgroups, each of index p in the preceding, whose intersection is trivial.

Therefore the generalization of Theorem 10 is:

THEOREM 11. *An abelian group is in the class C_r if and only if it has no nontrivial element whose order is relatively prime to r.*

I am grateful to Charles Weibel and Ali Nesin for contributing ideas which are crucial to the analysis and to David Hecker for showing the problem to me.

REFERENCE

1. A. P. Hillman, The William Lowell Putnam Mathematics Competition, this MONTHLY, 81 (December, 1974).

The Wedderburn Theorem of Finite Division Rings

JOHN SCHUE
Department of Mathematics and Computer Science, Macalester College, St. Paul, MN 55105

Last year, a senior honors student and I, while working on properties of finite rings, went through the two proofs, given in [**1**] of the Wedderburn theorem, that a finite division ring is a field. The first proof is a variation (due to Witt) of the original proof in 1905, and is largely number theoretic, while the second is essentially that given by Herstein in [**2**]. It is much more algebraic in nature (intentionally so) but becomes rather complicated technically. The proof presented here is in the same spirit but more conceptual than computational. It is intended to be accessible to the dedicated reader of a book like [**1**], with its major claim to acceptance perhaps resting on its drawing together a number of seemingly unrelated ideas developed earlier in the book. There are, of course, numerous other proofs available in the literature. Some, such as those found in [**4**, p. 225] (where it is an exercise) or in [**3**] obtain the theorem as a specialization of the much more general theory of simple algebras, while another, almost entirely group theoretic, can be found in [**5**].

Following Herstein, we let D be a finite division ring with center Z and assume that any division ring of smaller order than D is a field. We assume $Z \neq D$, and will show this leads to a contradiction. For $a \in d$, if $C(a) = \{x \in D\colon ax = xa\}$, then $C(a)$ is a subdivision ring containing Z and a so that, if $a \notin Z$, $C(a)$ is proper and hence is a field. Since any proper subdivision ring containing a also is contained in $C(a)$, then $C(a)$ is maximal in D as a subdivision ring. We let D^* be the multiplicative group of nonzero elements of D and, for a subset S of D, $S^* = S \cap D^*$. If p is the characteristic of D then p is a prime integer and $|D| = p^n$ for some integer n. For a, $x \in D$ we let $D_a x = [a, x] = ax - xa$. Then D_a is an additive operator of D with kernel $C(a)$ and, by Lemma 7.2.1, p. 369 of [**1**], we have $D_a^p = D_{ap}$.

Suppose now b is chosen as a fixed element of D with $b \notin Z$ and $K = C(b)$. Then K is a finite field of order p^m for some $m < n$. If $N^* = \{x \in D^*: xKx^{-1} \subset$

$K\}$, then N^* is a subgroup of D^* containing K^* as a normal subgroup. Since K^* is a cyclic group, we can choose a generator g for K^*. Then, for $x \in D^*$, $x \in N^*$ is equivalent to $xgx^{-1} \in K$ and, for $y \in N^*$, we will have $y \in Kx$ if and only if $xgx^{-1} = ygy^{-1}$. We let $q = |N^*/K^*|$.

Now D is a (left) vector space over K and, for $a \in K$, D_a is a K-linear operator on D. From the above, we have

$$D_a^{p^m} = D_a^{p^m} = D_a.$$

Thus the minimal polynomial for D_a is a divisor of

$$\lambda^{p^m} - \lambda$$

and necessarily splits in $K[\lambda]$ into distinct first degree factors so that D_a is diagonalizable. Applying this to D_g, for $E_k = \{x\colon [g, x] = kx\}$, we have $E_0 = K$ and $D = \Sigma E_k$, where the sum is direct and taken over all $k \in K$ with $E_k \neq 0$. Now, if $x \in D^*$, $gx - xg = kx$ for some $k \in K$ is equivalent to requiring that x belong to N^*. Moreover, $y \in E_k$ is equivalent to $y \in Kx$. Then each E_k is a K-subspace of dimension 1 and E_k^* is the coset K^*x in N^*. Hence, $\dim_K D = q$.

From the structure of finite fields it follows readily that K is a Galois extension of Z. We can identify N^*/K^* with a subgroup of $G(K/Z)$ and, if J is the fixed field for N^*/K^*, $a \in J$ implies $xax^{-1} = a$ for all $x \in N^*$. Then D_a is zero on each E_k so that $D_a = 0$ and $a \in Z$. Hence $N^*/K^* = G(K/Z)$, implying $\dim_Z K = |N^*/K^*| = q$.

Combining the results above leads to $\dim_Z D = (\dim_K D)(\dim_Z K) = q^2$ and $\dim_Z C(b) = q$ for all $b \in D$, $b \notin Z$. If $r = |Z|$ then

$$|D^*| = r^{q^2} - 1$$

and $|C(b)^*| = r^q - 1$ for $b \notin Z$. Thus, if s is the number of conjugacy classes containing more than one element, the class equation applied to D^* gives

$$r^{q^2} - 1 = (r-1) + s\left(r^{q^2} - 1/r^q - 1\right),$$

which implies $r^{q(q-1)} + \cdots + r^q + 1$ must be a divisor of $r - 1$, giving the desired contradiction.

REFERENCES

1. I. N. Herstein Topics in Algebra, 2nd ed., Xerox, 1975.
2. ______, Webberburn's theorem and a theorem of Jacobson, this MONTHLY, 68 (1961) 249–251.
3. ______, Non-Commutative Rings, MAA Carus Monograph No. 15, John Wiley and Sons, (1968).
4. N. Jacobson, Basic Algebra II, W. S. Freeman, 1980.
5. T. J. Kaczynski, Another proof of Wedderburn's theorem, this MONTHLY, 71 (1964) 652–653.

Common Hyperplane Medians for Random Vectors

THEODORE P. HILL*
School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332

Suppose some black and some white points are sprinkled randomly on the plane (e.g., salt and pepper on a table). Is there always a straight line so that at least half

*Partially supported by NSF Grants DMS-84-01604 and DMS-86-01608.

of the black points and at least half of the white points lie on one side of the line (including the line), and at least half of each color lies on the other side (again including the line)? Steinhaus' "Ham Sandwich Theorem" (Corollary 1 below) says there is always a line bisecting the total *areas* occupied by both the black and the white spots, but does not answer the question for true (volumeless) points. The purpose of this note is to prove a generalization of the Ham Sandwich Theorem which includes *all* distributions: continuous, purely atomic, and mixed. If the black and white points represent statistical data or observations, then the positions where they fall are *random vectors*, and it is in this terminology that the theorem will be stated. (The two-dimensional example of black and white points on a plane was used only for illustrative purposes; what is essential is that there are no more than n colors in Euclidean n-space.)

A *median* of a random variable is a real number α with the property that the probability the random variable is less than or equal to α, and the probability that it is bigger than or equal to α, are both at least one-half. One way of generalizing the notion of a median of a random variable to that of a median of a random vector is via hyperplanes.

DEFINITION. A *hyperplane median* for an n-dimensional random vector $\vec{X} = (X_1, \ldots, X_n)$ is an $(n-1)$-dimensional hyperplane

$$\sum_{j=1}^{n} a_j x_j = b \text{ in } \mathbb{R}^n \text{ satisfying}$$

$$P\left(\sum_{j=1}^{n} a_j X_j \geqslant b\right) \geqslant 1/2 \text{ and } P\left(\sum_{j=1}^{n} a_j X_j \leqslant b\right) \geqslant 1/2.$$

Clearly every finite-dimensional random vector has a hyperplane median, and the main purpose of this note is to prove the following generalization of this fact.

THEOREM 1. *Every collection of $m \leq n$ random n-dimensional vectors has a common hyperplane median.*

An immediate corollary is the Ham Sandwich Theorem, apparently first proved by Steinhaus (cf. **[2]**, p. 291).

COROLLARY 1. (Ham Sandwich Theorem) *If $B_1, \ldots, B_n$ are bounded Lebesgue measurable sets in $\mathbb{R}^n$, then there is a hyperplane which simultaneously bisects the measure (n-dimensional volume) of each B_i.*

(The picturesque name Ham Sandwich Problem is also attributed to Steinhaus, who viewed it as the problem of simultaneously bisecting the ham, cheese, and bread in an ordinary ham sandwich by a single planar cut of a knife.)

Proof of Corollary. Without loss of generality, assume each B_i has strictly positive measure, and let $\vec{Y}_i$ be the n-dimensional random vector which is uniformly distributed on B_i. Then apply Theorem 1.1 to $\vec{Y}_1, \ldots, \vec{Y}_n$. □

The key idea in the proof of Theorem 1 is an application of the following theorem of Borsuk and Ulam **[1]**; unlike the case in the Ham Sandwich Theorem, however, the measures (distributions) involved are not assumed to be continuous (or even nonatomic), and the usual continuity argument fails.

BORSUK-ULAM THEOREM. *If f is a continuous map of the surface of the unit ball in $(n+1)$-dimensional space into n-dimensional space such that $f(-\vec{x}) = -f(\vec{x})$ for every $\vec{x}$, then there is some point on the ball mapped into the origin.*

Let $\vec{X}$ be an n-dimensional random vector with distribution μ (i.e., $\mu(A) = P(\vec{X} \in A)$ for every n-dimensional Borel set A), let $\vec{x} \cdot \vec{y}$ denote the usual inner product of $\vec{x}$ and $\vec{y}$; and let $\|\vec{x}\|$ be the norm of $\vec{x}$.

DEFINITION. For $\vec{v} \in \mathbb{R}^n$, $m_*(\vec{v})$, $m^*(\vec{v})$, and $m(\vec{v})$ are the lower, upper, and *midpoint* (respectively) *medians* for $\vec{X}$ orthogonal to $\vec{v}$. That is, m_*, m^*, and m are the functions from $\mathbb{R}^n \to \mathbb{R}$ given by:

$$m_*(\vec{v}) = \inf\left\{a\colon P\left(\vec{X} \cdot \vec{v} \leqslant a\right) \geqslant 1/2 \quad \text{and} \quad P\left(\vec{X} \cdot \vec{v} \geqslant a\right) \geqslant 1/2\right\};$$

$$m^*(\vec{v}) = \sup\left\{a\colon P\left(\vec{X} \cdot \vec{v} \leqslant a\right) \geqslant 1/2 \quad \text{and} \quad P\left(\vec{X} \cdot \vec{v} \geqslant a\right) \geqslant 1/2\right\};$$

and

$$m(\vec{v}) = \left(m_*(\vec{v}) + m^*(\vec{v})\right)/2.$$

The proof of Theorem 1 will be based on two lemmas: the first lists some easy facts concerning m_*, m^* and m; and the second is the key continuity result.

LEMMA 1.
(i) *$m_*(\vec{v})$, $m^*(\vec{v})$, and $m(\vec{v})$ are finite and attained for all $\vec{v}$;*
(ii) *$m_*(\vec{0}) = m^*(\vec{0}) = m(\vec{0}) = 0$;*
(iii) *$m(-\vec{v}) = -m(\vec{v})$; and*
(iv) *if $\vec{v} \neq \vec{0}$ and $m(\vec{v}) = \alpha$, then $\vec{x} \cdot \vec{v} = \alpha$ is a hyperplane median for $\vec{X}$.*

LEMMA 2. *If the support of μ is compact, then m_*, m^*, and m are continuous.*

Proof. Only the argument for the continuity of m_* will be given; that for m^* follows similarly, and together they imply the continuity of m.

Let $K =$ support of μ; since K is compact there exists an $M > 0$ such that $\|\vec{x}\| \leqslant M$ for all $\vec{x}$ in K.

Fix $\vec{u} \in \mathbb{R}^n$ and $\varepsilon > 0$; then for every $\vec{v}$ in $\mathbb{R}^n$ satisfying $\|\vec{u} - \vec{v}\| < \varepsilon/M$,

$$|\vec{x} \cdot (\vec{u} - \vec{v})| < \varepsilon \text{ for all } \vec{x} \text{ in } K. \tag{1}$$

Let $\alpha = m_*(\vec{u})$. Then

$$\begin{aligned}\left\{\vec{x} \in K\colon \vec{x} \cdot \vec{v} \leqslant \alpha - 2\varepsilon\right\} &\subseteq \left\{\vec{x} \in K\colon \vec{x} \cdot \vec{u} \leqslant \alpha - \varepsilon\right\} \\ &\subseteq \left\{\vec{x} \in K\colon \vec{x} \cdot \vec{u} \leqslant \alpha\right\} \\ &\subseteq \left\{\vec{x} \in K\colon \vec{x} \cdot \vec{v} \leqslant \alpha + \varepsilon\right\}.\end{aligned} \tag{2}$$

(For the first inclusion in (2), observe that $\vec{x} \cdot \vec{v} \leqslant \alpha - 2\varepsilon$ implies $\vec{x} \cdot \vec{u} \leqslant \alpha - 2\varepsilon + \vec{x} \cdot (\vec{u} - \vec{v})$ which by (1) implies $\vec{x} \cdot \vec{u} \leqslant \alpha - \varepsilon$. The second inclusion is trivial and the third is analogous to the first.)

Since $\mu(K) = 1$, it follows from (2) and the definition of α that

$$\begin{aligned}P\left(\vec{X} \cdot \vec{v} \leqslant \alpha - 2\varepsilon\right) &\leqslant P\left(\vec{X} \cdot \vec{u} \leqslant \alpha - \varepsilon\right) < 1/2 \\ &\leqslant P\left(\vec{X} \cdot \vec{u} \leqslant \alpha\right) \leqslant P\left(\vec{X} \cdot \vec{v} \leqslant \alpha + \varepsilon\right).\end{aligned} \tag{3}$$

From (3) it follows that $\alpha - 2\varepsilon \leqslant m_*(\vec{v}) \leqslant \alpha + \varepsilon$, so since ε was arbitrary and $\alpha = m_*(\vec{u})$, this establishes the continuity of m_*. □

The conclusion of Lemma 2 may fail if the support of μ is not compact, even if μ is continuous (i.e., absolutely continuous with respect to Lebesgue measure).

Example 1. Suppose $\vec{X}$ is a continuous 2-dimensional random vector with strictly positive density of total mass $1/2$ on $A = \{(x,y)\colon 1 \leqslant x \leqslant 2,\ y \in \mathbb{R}\}$ and uniformly distributed with total mass $1/2$ on $B = \{(x,y)\colon 3 \leqslant x \leqslant 4,\ -1 \leqslant y \leqslant 1\}$. It is easy to check that $m((1,0)) = 5/2$, but that $\lim_{n\to\infty} m((1,n^{-1})) = 3$.

Proof of Theorem 1. Fix $m \le n$, let $\vec{X}_1,\dots,\vec{X}_m$ be n-dimensional random vectors with distributions $\mu_1,\dots,\mu_m$ respectively, and let m_i denote the midpoint-median function for $\vec{X}_i$ (as in Definition 2.1). Without loss of generality, $m = n$.

Case 1. The support of each μ_i, $i = 1,\dots,n$, is compact.

Let f be the function from the surface of the unit ball in $\mathbb{R}^{n+1}$ into $\mathbb{R}^n$ given by $f = (f_1,\dots,f_n)$, where $f_i(\vec{v},\beta) = m_i(\vec{v}) - \beta$ for all $i = 1,\dots,n$.

From Lemma 2 it follows easily that f is continuous, and from Lemma 1(iii) that $f(-\vec{v},-\beta) = -f(\vec{v},\beta)$, so the Borsuk-Ulam theorem implies there exists a point $(\vec{v}_0,\beta_0)$ on the surface of the unit ball in $\mathbb{R}^{n+1}$ satisfying $f(\vec{v}_0,\beta_0) = \vec{0}$, that is,

$$m_i(\vec{v}_0) = \beta_0 \text{ for all } i = 1,\dots,n. \tag{4}$$

By Lemma 1(ii) and the fact that $\|(\vec{v}_0,\beta_0)\| = 1$ it follows that $\vec{v}_0 \neq 0$, so the existence of a common median for $\vec{X}_1,\dots,\vec{X}_n$ follows by (4) and Lemma 1(iv). This completes the argument for Case 1.

General Case. Let $r > 0$ be such that $\mu_i(\vec{x}\colon \|\vec{x}\| \leqslant r) > 0$ for all $i = 1,\dots,n$. For $j = 1,2,\dots,$ let K_j denote the closed sphere $\{\vec{x}\colon \|\vec{x}\| \le r + j\}$ in $\mathbb{R}^n$, and let μ_{ij} denote the (re-normalized) restriction of μ_i to K_j, for $i = 1,\dots,n$. By Case 1, for each $j \geqslant 1$ there is a common median hyperplane $H_j = \{\vec{x} \in \mathbb{R}^n\colon \vec{a}_j \cdot \vec{x} = b_j\}$ for $\mu_{1j},\dots,\mu_{nj}$. Since $\mu_i(K_j) \to 1$ as $j \to \infty$, for all $i = 1,\dots,n$, it follows easily that a subsequence $\{H_{j_k}\}$ of $\{H_j\}$ converges (that is, there is an $\vec{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$ so that $\vec{a}_{j_k} \to \vec{a}$ and $b_{j_k} \to b$), and that the limiting hyperplane $H = \{\vec{x}\colon \vec{a}\cdot\vec{x} = b\}$ is a common median for $\vec{X}_1,\dots,\vec{X}_n$. □

A generalization of the notion of "median" that is often of use in statistics is that of a *β-quantile*, which is a number α with the property that the probability the random variable is $\leqslant \alpha$ is at least β, and the probability that it is $\geqslant \alpha$ is at least $1 - \beta$. Clearly every random variable has at least one β-quantile for each $\beta \in (0,1)$, and it is natural to ask whether Theorem 1 can be generalized to conclude that every collection of $m \leqslant n$ random variables has a common hyperplane β-quantile. The next example shows that in general the *only* common hyperplane β-quantiles which exist are for $\beta = 1/2$, i.e., medians.

Example 2. Let $\vec{X}_1$ be the 2-dimensional random vector which is uniformly distributed on the boundary of the unit circle in $\mathbb{R}^2$, and let $\vec{X}_2$ be the constant 2-dimensional random vector which is always the origin. It is easy to see that there are no common hyperplane β-quantiles for $\vec{X}_1$ and $\vec{X}_2$ for *any* $\beta \in (0,1)$ except $\beta = 1/2$.

(Note that for a fixed $\beta \in (0,1)$, $\beta \neq 1/2$, the above example can be modified slightly to yield *continuous* distributions with no common hyperplane β-quantile).

As pointed out by Carl Spruill, Theorem 1 can also be proved *from* the Ham Sandwich Theorem by adding $N(0,n^{-1})$ random variables to each component to

make each $\vec{X}_i$ (and hence each μ_i) continuous, applying for each n a slightly generalized version of the Ham Sandwich Theorem (for arbitrary continuous probability distributions), and then appealing to standard results for weak convergence of probability measures. The proof presented above has the advantage that it is more elementary (since it relies essentially only on the Borsuk-Ulam Theorem), and it also shows how to modify directly the classical proof of the Ham Sandwich Theorem to include noncontinuous distributions.

It should also be emphasized that both proofs of Theorem 1 (the given proof, and Spruill's proof) are highly nonconstructive, and it would perhaps be of interest to devise a practical, constructive method for determining the common hyperplane medians guaranteed by the theorem. Spruill has also raised the question of which curves or manifolds other than straight lines or hyperplanes can serve as common medians.

REFERENCES

1. K. Borsuk, Drei Sätze über die n-dimensionale euklidische Sphäre, *Fund. Math.*, 20 (1933) 177–190.
2. M. Agoston, Algebraic Topology, Marcel Dekker, New York, 1976.

EDITED BY JOAN P. HUTCHINSON AND STAN WAGON

Quadratic Reciprocity: Its Conjecture and Application

DAVID A. COX
Department of Mathematics, Amherst College, Amherst, MA 01002

If p and q are distinct odd primes, then the law of quadratic reciprocity states that

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{(p-1)(q-1)/4} \tag{0.1}$$

where (a/p) is the Legendre symbol, defined to be $+1$ when a is a quadratic residue modulo p and to be -1 otherwise. Quadratic reciprocity is one of the gems of elementary number theory, but at the same time, it's never been easy to see where the theorem comes from. Other formulations are not a great help. For example, statement (0.1) is equivalent to the following:

> If p or q is congruent to 1 modulo 4, then p is a quadratic residue modulo q if and only if q is a quadratic residue modulo p. If p and q are congruent to 3 modulo 4, then p is a quadratic residue modulo q if and only if q is a quadratic nonresidue modulo p. (0.2)

(Again, p and q are odd primes.) This version of the theorem is easier to read, but it still sheds no light on how to conjecture such an amazing result.

Of course, once we know to look for some relation between quadratic residues modulo different primes, it's fairly easy to find examples that lead to a statement like (0.1) or (0.2). But how does one begin to suspect that such a relation should exist?

This paper has two goals. The first is to answer the question just raised, i.e., to present a series of questions and examples that lead naturally to quadratic reciprocity. The second goal is to use quadratic reciprocity to prove the following theorems of Fermat for odd primes p:

$$\begin{aligned}
p &= x^2 + y^2, & x, y &\in \mathbb{Z} & &\Leftrightarrow & p &\equiv 1 \bmod 4\\
p &= x^2 + 2y^2, & x, y &\in \mathbb{Z} & &\Leftrightarrow & p &\equiv 1, 3 \bmod 8\\
p &= x^2 + 3y^2, & x, y &\in \mathbb{Z} & &\Leftrightarrow & p &= 3 \text{ or } p \equiv 1 \bmod 3.
\end{aligned} \tag{0.3}$$

Our two goals are closely related, for we will use Fermat's results to motivate the questions that lead to reciprocity. This might seem circular, but it's how things happened historically. Euler knew of Fermat's theorems, but since Fermat never published the proofs, Euler had to find his own. This took many years, and along the way Euler discovered quadratic residues and quadratic reciprocity. This paper will follow Euler closely, both in the examples leading to reciprocity and in the proofs of (0.3). For an excellent account of Euler's work on number theory, the reader should consult Weil's book **[4]**.

There is one other aspect of our second goal which deserves mention. Many books on number theory present quadratic reciprocity but fail to give interesting

applications. It is important to show that quadratic reciprocity is not an isolated result, but rather is a theorem that leads immediately to other theorems of independent interest. (I should point out that several books do this very nicely, in particular Ireland-Rosen **[2]** and Nagell **[3]**. I am grateful to the referees for pointing out these references and for suggesting improvements in the exposition.)

1. We will first discuss Euler's strategy for proving Fermat's results (0.3). One half of the theorem is easy: if $p = x^2 + ny^2$, $n = 1$, 2 or 3, then the desired congruence for p follows by considering x and y modulo 4, 8 and 3 respectively. The converse is more difficult, and this is where Euler's strategy comes in. For the case of $x^2 + y^2$, it consists of the following two steps:

Step 1. If $p \equiv 1 \bmod 4$ is prime, then show that p divides a sum of the form $x^2 + y^2$, where x and y are relatively prime integers.

Step 2. If an odd prime p divides a sum of the form $x^2 + y^2$, where x and y are relatively prime, then show that p is a sum of two squares.

The cases $x^2 + 2y^2$ and $x^2 + 3y^2$ can be formulated similarly, and it is clear that these steps suffice to prove Fermat's results (0.3).

Notice that both steps involve prime divisors of $x^2 + ny^2$, where $n = 1, 2$ or 3. It thus makes sense to look more closely at this situation, and as happens often in mathematics, we can better understand our problem by posing it more generally. Hence, in considering prime divisors of $x^2 + ny^2$, we should let n be an arbitrary integer, positive or negative. We still want x and y to be relatively prime, for otherwise any prime could divide $x^2 + ny^2$—just take x and y to be multiples of p.

We can restate our problem in terms of the Legendre symbol as follows:

LEMMA 1.1. *Let p be a prime not dividing n. Then there are relatively prime integers x and y such that $p | x^2 + ny^2$ if and only if $(-n/p) = 1$.*

Proof. Suppose that p divides such a number $x^2 + ny^2$. Then $x^2 \equiv -ny^2 \bmod p$. Since x and y are relatively prime, it follows that $p \nmid y$. The integers modulo p form a field, so that $yb \equiv 1 \bmod p$ for some b. Multiplying our congruence by b^2, we see that $(xb)^2 \equiv -n \bmod p$, which implies that $(-n/p) = 1$. The converse is trivial, and the lemma is proved.

We are now ready to turn to Step 1. Using Lemma 1.1, we see that is suffices to prove the following for primes p:

$$\begin{aligned}
\left(\frac{-1}{p}\right) = 1 \quad &\Leftrightarrow \quad p \equiv 1 \bmod 4\\
\left(\frac{-2}{p}\right) = 1 \quad &\Leftrightarrow \quad p \equiv 1, 3 \bmod 8\\
\left(\frac{-3}{p}\right) = 1 \quad &\Leftrightarrow \quad p \equiv 1 \bmod 3.
\end{aligned} \tag{1.2}$$

As we did before, let's pose this more generally. Thus we want congruence conditions on p that imply $(-n/p) = 1$. Looking at (1.2), we see that the way to unify the congruence conditions is to work modulo $4n$.

It's now time to do some examples. We will work out the case $n = -5$ in detail. This means studying the remainders modulo 20 of those primes p for which $(5/p) = 1$. One way to generate such primes is to factor the numbers $x^2 - 5$ for variable x. If we let x range from 1 to 30, we get the following prime factors (excluding 2 and 5):

$$11, 19, 29, 31, 41, 59, 61, 71, 79, 89$$

$$109, 131, 139, 179, 181, 191, 251, 471, 571.$$

Working modulo 20, all of these primes are congruent to ± 1 or ± 11, and the first row lists all primes under 100 which satisfy this congruence condition. If we let x range from 1 to 80, we get all primes under 200 that satisfy the congruence. It appears that we will eventually pick up all such primes, which leads to the conjecture that for a prime $p \neq 2, 5$,

$$\left(\frac{5}{p}\right) = 1 \quad \Leftrightarrow \quad p \equiv \pm 1, \pm 11 \bmod 20.$$

Other examples can be worked out similarly. Here are some sample cases:

$$\begin{aligned}
\left(\frac{-3}{p}\right) = 1 \quad &\Leftrightarrow \quad p \equiv 1, 7 \bmod 12\\
\left(\frac{-5}{p}\right) = 1 \quad &\Leftrightarrow \quad p \equiv 1, 3, 7, 9 \bmod 20\\
\left(\frac{-7}{p}\right) = 1 \quad &\Leftrightarrow \quad p \equiv 1, 9, 11, 15, 23, 25 \bmod 28 \qquad (1.3)\\
\left(\frac{3}{p}\right) = 1 \quad &\Leftrightarrow \quad p \equiv \pm 1 \bmod 12\\
\left(\frac{5}{p}\right) = 1 \quad &\Leftrightarrow \quad p \equiv \pm 1, \pm 11 \bmod 20\\
\left(\frac{7}{p}\right) = 1 \quad &\Leftrightarrow \quad p \equiv \pm 1, \pm 3, \pm 9 \bmod 28.
\end{aligned}$$

Here we always assume that p is an odd prime not dividing n.

In looking for something to unify these conjectures, the bottom three results of (1.3) offer the most hope because of the $\pm$'s. We will thus concentrate on the case (N/p) for $N > 0$.

The key problem is to discern a pattern in the numbers that follow the $\pm$'s. Of course, numbers in congruences can be written many ways. For example, $11 \equiv -9 \bmod 20$ and $3 \equiv -25 \bmod 28$. Using these two facts we can rewrite the bottom part of (1.3) as follows:

$$\begin{aligned}
\left(\frac{3}{p}\right) = 1 \quad &\Leftrightarrow \quad p \equiv \pm 1 \bmod 12\\
\left(\frac{5}{p}\right) = 1 \quad &\Leftrightarrow \quad p \equiv \pm 1, \pm 9 \bmod 20 \qquad (1.4)\\
\left(\frac{7}{p}\right) = 1 \quad &\Leftrightarrow \quad p \equiv \pm 1, \pm 25, \pm 9 \bmod 28.
\end{aligned}$$

All of a sudden there is a pattern: all of the numbers that appear are odd squares!
Before we get too excited, let's try another case:

$$\left(\frac{6}{p}\right) = 1 \quad \Leftrightarrow \quad p \equiv \pm 1, \pm 5 \bmod 24.$$

Unfortunately, ± 5 is *not* congruent to a square modulo 24. The same phenomenon happens for $(10/p)$ and $(14/p)$. So why does the pattern work for 3, 5 and 7 but not for 6, 10 and 14? The obvious difference is that the former numbers are all prime!
This says that there is something very special about the prime case, i.e., the case of $(q/p) = 1$, where p and q are distinct odd primes. From (1.4), we get the following conjecture.

CONJECTURE 1.5. *If p and q are distinct odd primes, then*

$$\left(\frac{q}{p}\right) = 1 \Leftrightarrow p \equiv \pm\beta^2 \bmod 4q \text{ for some odd } \beta.$$

Notice that we have now given an answer to the basic question posed in the introduction—the above examples single out the prime case very nicely. It remains to show that Conjecture 1.5 is equivalent to the usual formulation of quadratic reciprocity. We do this as follows.

THEOREM 1.6. *Conjecture* 1.5 *is equivalent to the law of quadratic reciprocity for distinct odd primes* (*see* (0.1)).

Proof. Let p and q be distinct odd primes, and set $p^* = (-1)^{(p-1)/2}p$. Note that $p^* \equiv 1 \bmod 4$. We will assume the following two properties of the Legendre symbol:

$$\begin{gathered}\left(\frac{-1}{q}\right) = (-1)^{(q-1)/2} \\ \left(\frac{ab}{q}\right) = \left(\frac{a}{q}\right)\left(\frac{b}{q}\right)\end{gathered} \tag{1.7}$$

(see [3, p. 135]). Using (1.7) it is an easy exercise to show that quadratic reciprocity is equivalent to the statement

$$\left(\frac{q}{p}\right) = \left(\frac{p^*}{q}\right).$$

Since each side equals ± 1, it follows that quadratic reciprocity can be written as the equivalence

$$\left(\frac{q}{p}\right) = 1 \Leftrightarrow \left(\frac{p^*}{q}\right) = 1,$$

and comparing this to Conjecture 1.5, it thus suffices to show

$$p \equiv \pm\beta^2 \bmod 4q \Leftrightarrow \left(\frac{p^*}{q}\right) = 1. \tag{1.8}$$

Note that $\beta^2 \equiv 1 \bmod 4$ since β is odd. Thus the $\pm$ sign in (1.8) must be

$(-1)^{(p-1)/2}$, and we then have

$$p \equiv \pm\beta^2 \bmod 4q \Leftrightarrow p \equiv (-1)^{(p-1)/2}\beta^2 \bmod 4q$$
$$\Leftrightarrow p^* \equiv \beta^2 \bmod 4q.$$

Now, to prove (1.8), suppose that $p^* \equiv \beta^2 \bmod 4q$. This implies $p^* \equiv \beta^2 \bmod q$, so that $(p^*/q) = 1$ follows immediately. Conversely, if $(p^*/q) = 1$, then $p^* \equiv \alpha^2 \bmod q$ for some α. Letting $\beta = \alpha$ or $\alpha + q$, depending on whether α is odd or even, we obtain $p^* \equiv \beta^2 \bmod 4q$, and the theorem is proved.

With this theorem and the preceding examples, it becomes clearer how someone could have conjectured quadratic reciprocity. In our case, however, the "someone" is none other than Euler, for we have followed roughly the same path he took to discover reciprocity. In fact, our examples (1.3) are taken from Euler's 1744 paper "Theoremata circa divisores numerorum in hac forma $pxx \pm qyy$ contentorum" [**1**, vol. 2, pp. 194–222], and in the same paper there are a series of Annotations that include Conjecture 1.5 as a special case. For more details, see [**4**, pp. 186–187, 204–210].

We should also mention the other parts of reciprocity, the so-called complementary theorems for odd primes p:

$$\left(\frac{-1}{p}\right) = (-1)^{(p-1)/2}$$
$$\left(\frac{2}{p}\right) = (-1)^{(p^2-1)/8}. \tag{1.9}$$

Notice that we've already used the formula for $(-1/p)$ in the proof of Theorem 1.6.

For our purposes, the key fact is that once we've done a full treatment of reciprocity (statements (0.1) and (1.9)), Step 1 of the proof of Fermat's results follows immediately—the equivalences (1.2) are now an easy exercise which we leave to the reader. It remains to prove Step 2, which will be done in §2.

2. Earlier, we stated Step 2 only for the case $x^2 + y^2$. We will state the general case, for $x^2 + ny^2$, $n = 1, 2, 3$, as the following theorem.

THEOREM 2.1. *Let p be an odd prime that divides a sum $a^2 + nb^2$, where $n = 1, 2$ or 3 and a and b are relatively prime. Then p can be written in the form $x^2 + ny^2$.*

Proof. We will follow the arguments Euler used in 1752 for the case $x^2 + y^2$ [**1**, vol. 2, pp. 300–307]. We will generalize his arguments to $x^2 + ny^2$, $n = 1, 2, 3$.

The crucial step is the following lemma.

LEMMA 2.2. *Let $q = x^2 + ny^2$, n a positive integer, and suppose that q divides a number $N = a^2 + nb^2$, where a and b are relatively prime. If either q is prime, or $q = 4$ and $n = 3$, then $N/q = c^2 + nd^2$, where c and d are relatively prime.*

Proof. Let us first consider the case where q is prime. Since q divides both $x^2N = x^2(a^2 + nb^2)$ and $a^2q = a^2(x^2 + ny^2)$, it divides their difference

$$x^2(a^2 + nb^2) - a^2(x^2 + ny^2) = n(x^2b^2 - a^2y^2) = n(xb - ay)(xb + ay).$$

Since q is prime, it must divide one of these factors.

If $q|n$, then $q = n$ since $q = x^2 + ny^2$. Hence $n|N = a^2 + nb^2$, so that $n|a$, i.e., $a = nd$. Then $N = n^2d^2 + nb^2$, which implies $N/q = b^2 + nd^2$, as desired.

If $q|xb - ay$ or $q|xb + ay$, we can assume that the former holds by changing the sign of y. Then $xb - ay = dq = d(x^2 + ny^2)$. This implies that

$$xb - dx^2 = ay + dny^2 = y(a + ndy), \tag{2.3}$$

from which we conclude that $x|y(a + ndy)$. Since x and y are relatively prime (q is prime), we must have $x|a + ndy$, i.e.,

$$a + ndy = cx \tag{2.4}$$

so that $a = cx - ndy$. Substituting (2.4) into (2.3), we obtain

$$x(b - dx) = y(cx),$$

which implies that $b = dx + cy$.

However, we also have the famous identity

$$(c^2 + nb^2)(x^2 + ny^2) = (cx - ndy)^2 + n(dx + cy)^2.$$

Using the above formulas for a and b, this becomes

$$(c^2 + nd^2)q = a^2 + nb^2 = N,$$

so that $N/q = c^2 + nd^2$, as desired. Since a and b are relatively prime, the formulas for a and b show that c and d are also relatively prime.

It remains to consider the case $n = 3$ and $q = 4$. Here, we have $4|a^2 + 3b^2$, so that a and b have the same parity. Since a and b are relatively prime, they must be odd. Since $4 = 1^2 + 3 \cdot 1^2$, the argument for the prime case (with $x = y = 1$) would work, provided that $4|b - a$ or $4|b + a$. But the latter holds for any pair of odd numbers, which proves the lemma in this case.

The proof of this lemma is similar in strategy to Lagrange's proof of the famous four square theorem, which asserts that every positive integer is a sum of four or fewer squares (see, for example, [**3**, Theorem 102]).

To complete the proof of Theorem 2.1, let p be an odd prime dividing $a^2 + nb^2$, where a and b are relatively prime. Assume that p itself is not of this form. We will show that there is an odd prime $q < p$ with exactly the same properties. We would then be done by Fermat's principle of infinite descent: applying the same argument to q would give us $q' < q$, and continuing we would get an infinite decreasing sequence $p > q > q' > \cdots$ of positive integers, which contradicts the well-ordering property.

To produce q, we work with $a^2 + nb^2$. It is divisible by p, and remains so if we replace a and b by $a - kp$ and $b - \ell p$ respectively. Furthermore, we may choose k and ℓ so that $|a - kp| < p/2$ and $|b - \ell p| < p/2$ because p is odd. Thus we may assume that $p|a^2 + nb^2$ where $|a| < p/2$ and $|b| < p/2$. Since $n \leqslant 3$, it follows that $a^2 + nb^2 < (p/2)^2 + 3(p/2)^2 = p^2$. Thus $a^2 + nb^2$ can be written as

$$a^2 + nb^2 = pq_1 \cdots q_r, \tag{2.5}$$

where the primes q_i all satisfy $q_i < p$.

We claim that one of these q_i's is odd and not of the form $x^2 + ny^2$. Assume not. Then all of the odd q_i's can be written as $x^2 + ny^2$, so that by repeatedly applying

Lemma 2.2 we can eliminate all of the odd q_i's from (2.5). This leaves us with

$$a^2 + nb^2 = 2^a p.$$

If $n = 1$ or 2, we can also apply Lemma 2.2 to $2 = 1^2 + 1^2 = 0^2 + 2 \cdot 1^2$ to eliminate factors of 2, showing that $p = a^2 + nb^2$, a contradiction. If $n = 3$, the case $q = 4$ of Lemma 2.2 shows that we can reduce to either $p = a^2 + 3b^2$ or $2p = a^2 + 3b^2$. It remains to show that the latter case cannot occur. But $p | a^2 + 3b^2$ implies $(-3/p) = 1$, which by quadratic reciprocity means $p \equiv 1 \bmod 3$, so that $2p \equiv 2 \bmod 3$. Yet $2p = a^2 + 3b^2$ implies $2p \equiv a^2 \equiv 1 \bmod 3$, and thus we have a contradiction.

This completes the proof of Theorem 2.1.

This finishes our proof of Fermat's three theorems (0.3), and the only gap is that we haven't proved quadratic reciprocity. So once a course in number theory covers reciprocity, there really are nice applications waiting to be proved, and some of them can even help motivate the statement of the theorem itself.

REFERENCES

1. L. Euler, Opera Omnia, series I, volume 2, Teubner, Leipzig and Berlin, 1915.
2. K. Ireland and M. Rosen, A Classical Introduction to Modern Number Theory, Springer-Verlag, New York-Berlin-Heidelberg, 1982.
3. T. Nagell, Introduction to Number Theory, Chelsea, New York, 1981.
4. A. Weil, Number Theory: An Approach through History, Birkhäuser, Boston-Basel-Stuttgart, 1984.

A Geometric Interpretation of the Riemann-Stieltjes Integral

GREGORY L. BULLOCK,
Mission Research Corporation, Santa Barbara, CA 93102

1. Introduction. For many students of mathematics, the moment of truth comes in a rigorous study of real analysis. A promising student may become discouraged during a real analysis course and subsequently abandon serious hopes of pursuing a degree in mathematics and may even dissuade other students from doing the same. A graphical explanation of some theorems of analysis can simplify the task of an instructor and may save some students from unwarranted distress.

Herein are presented geometric interpretations of the *Riemann-Stieltjes integral* (or, simply, the *Stieltjes integral*) and a few of the associated theorems as given in [**1**, chapter 6]. For convenience, the definition of the integral is reproduced, and the theorems that are interpreted here are stated without proof.

2. Definition of the Integral. Let $[a, b]$ be a given interval. Define a partition P of $[a, b]$ to be a set of points $x_0, x_1, \dots, x_n$, where $a = x_0 \le x_1 \le \cdots \le x_n = b$. Let α be a monotonically increasing function on $[a, b]$. For each partition P of $[a, b]$ write $\Delta\alpha_k = \alpha(x_k) - \alpha(x_{k-1})$.

For any real function f that is bounded on $[a, b]$ let $M_k = \max\{f(x), x_{k-1} \le x \le x_k\}$, $m_k = \min\{f(x), x_{k-1} \le x \le x_k\}$, and set

$$U(P, f, \alpha) = \sum_{k=1}^{n} M_k \Delta\alpha_k,$$

and

$$L(P, f, \alpha) = \sum_{k=1}^{n} m_k \Delta\alpha_k.$$

If there is a unique number I that satisfies the inequality $L(P, f, \alpha) \leq I \leq U(P, f, \alpha)$ for all partitions P of $[a, b]$, then I is called the *Stieltjes integral* of f from a to b and is denoted by

$$\int_a^b f\, d\alpha \quad \text{or} \quad \int_a^b f(x)\, d\alpha(x).$$

THEOREM 1. *Assume α increases monotonically and α' is Riemann integrable on $[a,b]$. Let f be a bounded, real, Riemann integrable function on $[a, b]$. Then*

$$\int_a^b f\, d\alpha = \int_a^b f(x)\alpha'(x)\, dx.$$

THEOREM 2. *If $a < s < b$, f is bounded on $[a, b]$, f is continuous at s, and*

$$\alpha(x) = \begin{cases} 0, & x \leq s \\ 1, & x > s, \end{cases}$$

then $\int_a^b f\, d\alpha = f(s)$.

3. In 3-D. When dependent sets are graphed, each set is assigned a dimension with a direction perpendicular to that of the dimensions of the other sets. So $f(x)$ and $\alpha(x)$ are graphed as in FIGURES 1 and 2. For this discussion, let $\alpha(x)$ be differentiable and observe that $\alpha(x)$ is monotonically increasing in agreement with the definition of the Stieltjes integral. If $f(x)$, $\alpha(x)$, and x are to be considered simultaneously, then each is assigned its own dimension and direction perpendicular to that of the other two. A key to finding a geometric interpretation of the integral is to notice that since $f(x)$ is independent of $\alpha(x)$, then for a given x, $f(x)$ must be constant with respect to $\alpha(x)$. So $f(x)$ is a cylinder (or more intuitively, a sheet) that is straight in the α-direction. If one looks along the sheet in the x-direction one may see hills and valleys, but one will see only flat terrain in the α-direction (see FIG. 3). The same result holds for $\alpha(x)$, which is independent of $f(x)$.

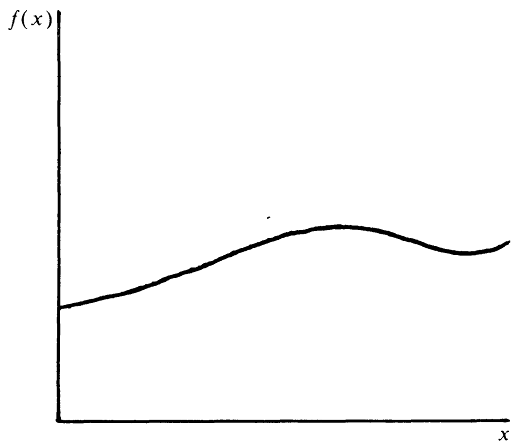


FIG. 1.

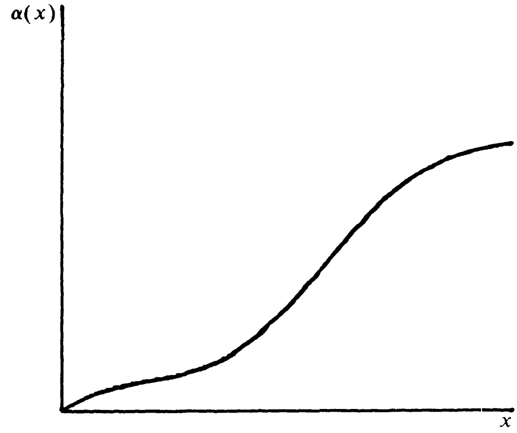


FIG. 2.

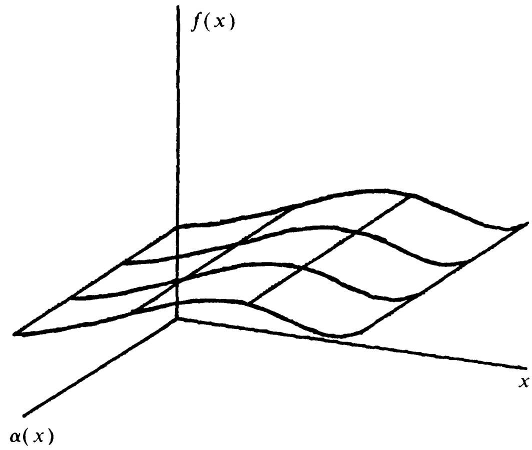


FIG. 3.

4. The Integrated Fence and its Illuminating Shadow. Assume still that $\alpha(x)$ is differentiable. If the α-x plane is thought of as horizontal and the f-direction as pointing straight up, then the surface to be considered is like a curved fence. The fence follows the curve traced by $\alpha(x)$, and the height of the fence is given by $f(x)$ (see FIG. 4). The fence is really the section of the α-sheet that is bounded between the $\alpha-x$ plane and the f-sheet. The Stieltjes integral integrates along this fence, summing the products of heights and infinitesimal widths. It uses f as the height in each interval, but for the differential width it considers only $\Delta\alpha_k$, the length of the infinitesimal interval in the α-direction. As a result, the area given by the integral is really the area of the projection of the fence onto the f-α plane (in the same sense that the "area under a curve" is defined as that given by the Riemann integral). If a spotlight with a beam parallel to the x-axis is positioned so as to aim toward the fence, then the area given by the integral is the area of the shadow of the fence on a wall built in the f-α plane (see FIG. 5).

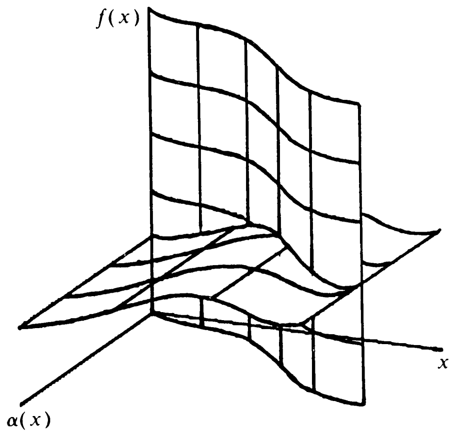


FIG. 4.

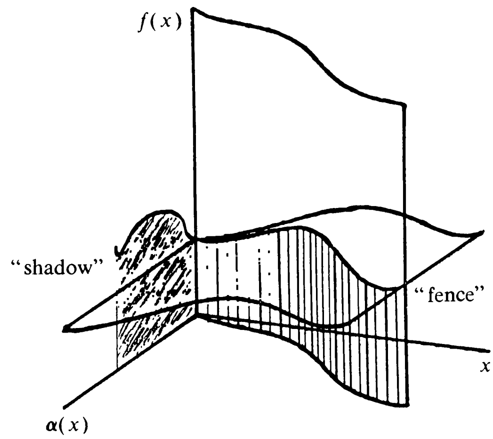


FIG. 5.

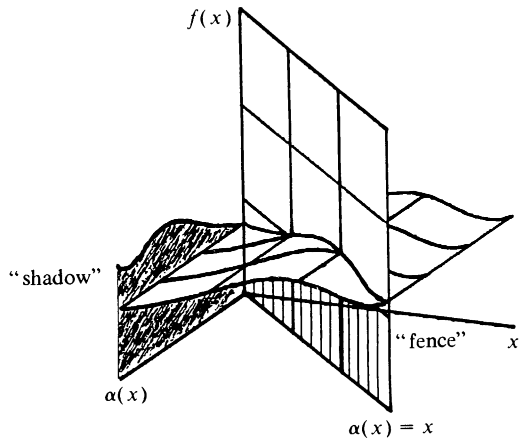


FIG. 6.

It is interesting to note that in this configuration, the Riemann integral is simply the area of the projection of the fence onto the f-x plane. It is easy to see that if $\alpha(x)$ is defined by $\alpha(x) = x$, then the fence allows a straight line from the origin that makes an angle of 45° with both the α-axis and the x-axis. In this case the area of the projection onto the f-α plane is equal to the area of the projection onto the f-x plane (see FIG. 6). Thus, the Stieltjes integral reduces to the Riemann integral. This can be verified by a simple application of Theorem 1.

5. Throwing a Curve. Consider the case where α is not a straight line. Define f as in FIGURE 7 (in three dimensions this is a plane), and let α_1 be given by FIGURE 8a and α_2 by FIGURE 9a. Let α_1 and α_2 be differentiable. The corresponding figures show the fences and their projections onto the f-α plane. A careful inspection of these shadows will convince the reader that the curve in α weights the area of the shadow. The values of x for which $\alpha(x)$ has the steepest slope correspond to regions

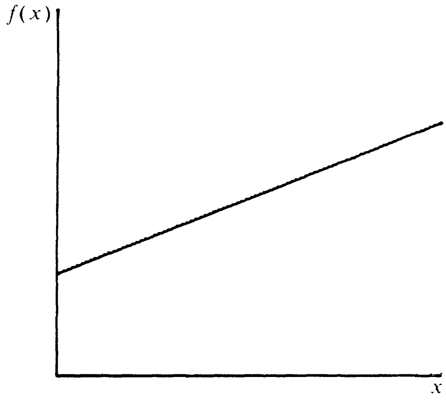


FIG. 7.

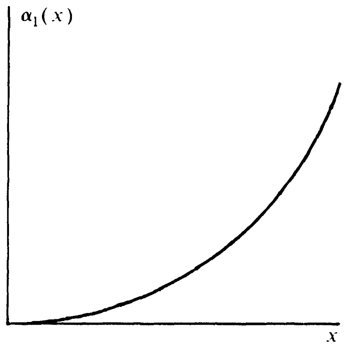


FIG. 8a.

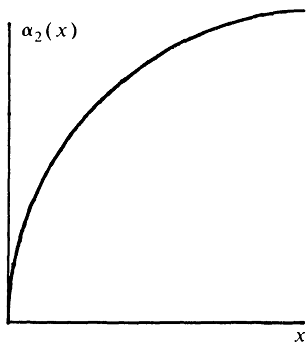


FIG. 8b.

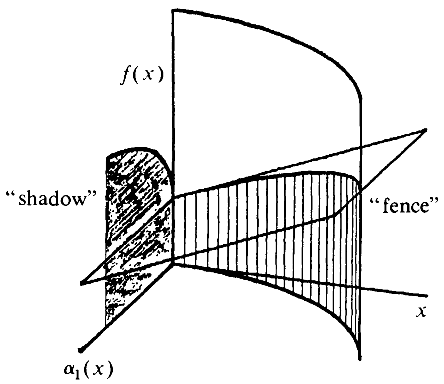


FIG. 9a.

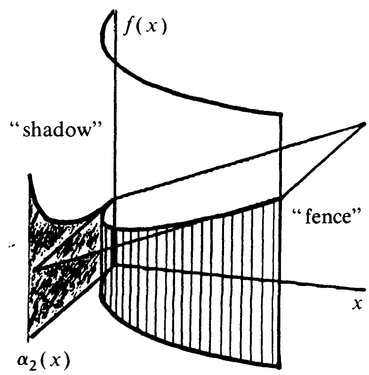


FIG. 9b.

of the fence that cast the most shadow and thereby carry the most weight in the integral. While trying to picture this, it may be useful to observe that regions of the fence in which $\alpha(x)$ has zero slope cast no shadow at all. This is the essence of the identity from Theorem 1

$$\int f d\alpha = \int f(x)\alpha'(x)\,dx \tag{1}$$

where $\alpha'(x) = d\alpha/dx$ is simply the weighting function.

6. A Gate in the Fence. Using the definition

$$\alpha(x) = \begin{cases} 0, & x \le s \\ 1, & x > s, \end{cases}$$

where f is continuous at s, the fence has a hole—or a gate—in it (see FIGS. 10 and 11). The width of the gate is unity, and the height of the fence at the gate is $f(s)$, so the area of the gate is $f(s)$. That the fence has no projection (other than two line segments) in the f-α plane only slightly complicates the analogy to the geometric interpretation of the integral. The complication can be resolved by simply filling in the fence at the hole, in other words, closing the gate door. The projection of the gate door has area $f(s)$ which is what the Stieltjes integral gives. This agrees with Theorem 2 above. This analogy is applicable whenever $\alpha(x)$ is monotonic and has at least one jump discontinuity.

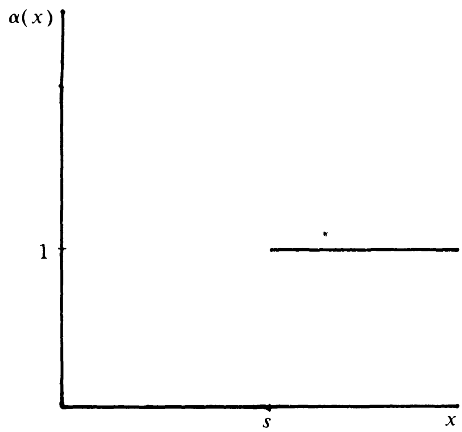


FIG. 10.

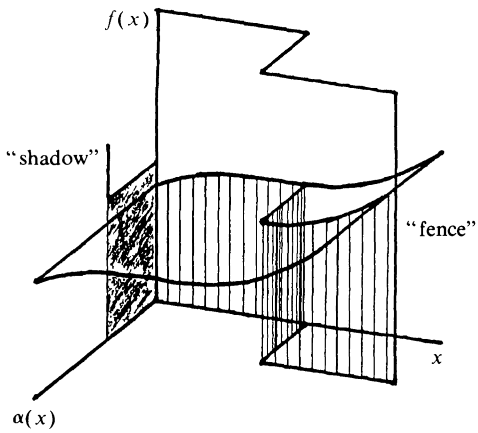


FIG. 11.

7. A Sidewise Glance. In the general definition of the Stieltjes integral, the restriction that $\alpha(x)$ be monotonically increasing is not imposed. For the following discussion, let the restriction be removed, and consider the case where $\alpha(x)$ is decreasing on some intervals. On intervals where $\alpha(x)$ is decreasing, $\alpha'(x)$ is negative, and the area of the corresponding projected region contributes negatively to the integral as suggested by equation (1). The analogy to the geometric interpreta-

tion of the integral may be applied directly by dividing the original interval into subintervals on which $\alpha(x)$ is monotonic. The shadow of each subinterval may be considered separately, and the area of each shadow may be added or subtracted (as determined by the sign of $\alpha'(x)$ on the subinterval) to determine the value of the integral as illustrated in FIGURES 12a–c. It is amusing to observe that in this configuration, the net region whose area is given by the Stieltjes integral typically assumes unusual shapes (see FIG. 13). In fact, the integral can give the sum of the areas of two (or more) regions connected only by a point as shown in FIGURES 14a-b.

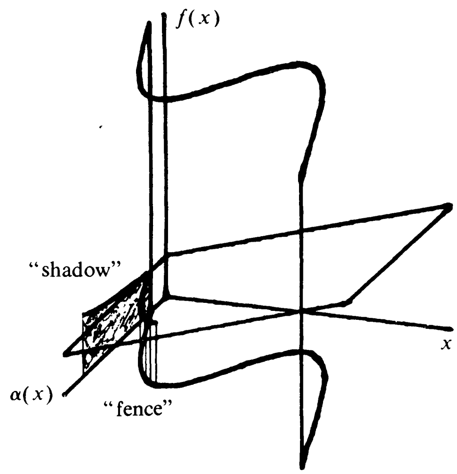


FIG. 12a.

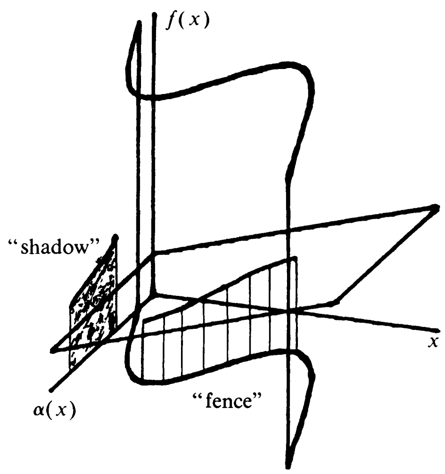


FIG. 12b.

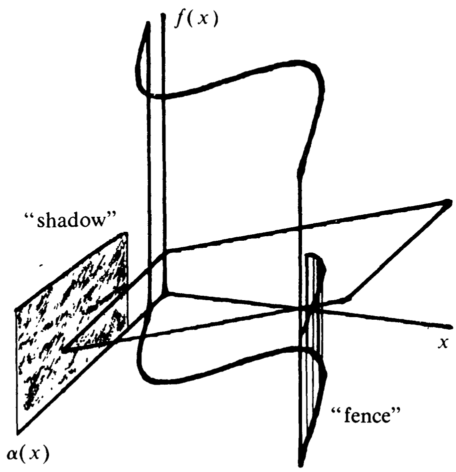


FIG. 12c.

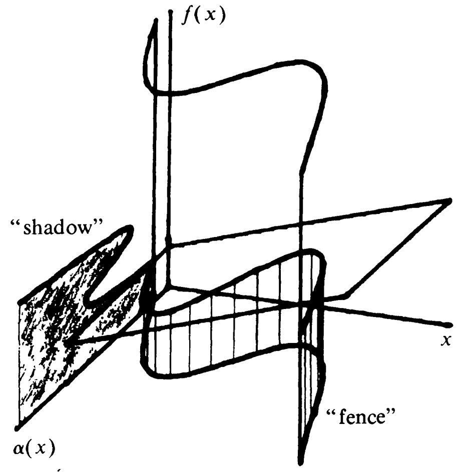


FIG. 13.

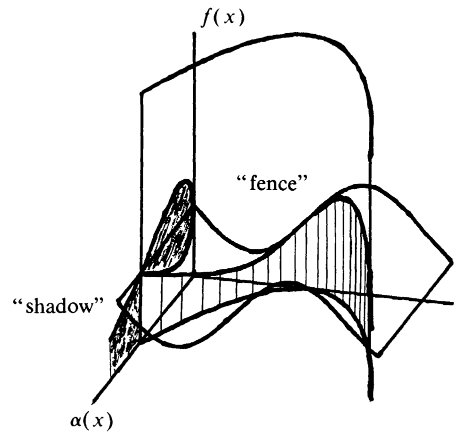


FIG. 14a.

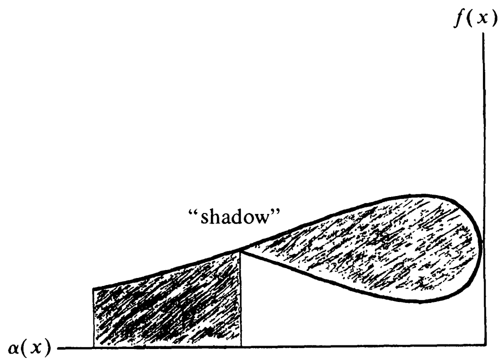


FIG. 14b.

8. Conclusion. Geometric interpretations of the Riemann-Stieltjes integral and several of its associated theorems have been presented. The author hopes that these will aid mathematics instructors in presenting difficult concepts associated with the integral.

I am indebted to Dick Epstein, my mathematics professor at the University of California, Berkeley, for his helpful suggestions and encouragement. I am grateful to Robert Drucker for his ever-illuminating observations.

REFERENCE

1. Walter Rudin, Principles of Mathematical Analysis, third edition, McGraw-Hill, New York, 1976.

PROBLEMS AND SOLUTIONS

For instructions about submitting ***proposed*** *problems for publication in this department see the inside front cover. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

Solutions *should be sent to the address given on the inside front cover. Two copies suffice.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

For instructions about submitting solutions of Problems, which should be mailed before September 30, 1988, see the inside front cover. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label if an acknowledgment is desired.

E 3265. *Proposed by D. M. Friedlen, Georgia Institute of Technology.*

Suppose n distinct pairs of socks are put into a laundry, where it is assumed that each of the $2n$ socks has precisely one mate. At the completion of the laundering operation, socks are drawn out one at a time. Suppose the first pair is realized on draw number T_n.

(a) Find an explicit formula for $E(T_n)$, the expected value of T_n, and determine

$$\lim_{n\to\infty} n^{-1/2}E(T_n).$$

(b) For positive x prove that

$$\lim_{n\to\infty} \mathbb{P}\{T_n \le xn^{1/2}\}$$

exists and find its value.

E 3266. *Proposed by John C. Turner, University of Waikato, Hamilton, New Zealand.*

Let N be the set of natural numbers. If $S \subseteq N$ and $n \in N$, let $S + n = \{s + n : s \in S\}$. Define a sequence of sets $S_1, S_2, S_3, \dots$ recursively by putting $S_1 = \{1\}$ and $S_k = (S_{k-1} + k) \cup \{2k - 1\}$ for $k = 2, 3, 4, \dots$. What is $N \setminus \bigcup_{k=1}^{\infty} S_k$?

E 3267. *Proposed by Barry Hayes, Donald Knuth, and Carlos Subi, Stanford University.*

Given a sequence $(x_1, x_2, \dots, x_l)$ of nonnegative integers in which $x_k > 1$ for some k, where $1 < k < l$, let us say that a "k-move" is the operation of replacing the subsequence (x_{k-1}, x_k, x_{k+1}) by $(x_{k-1} + 1, x_k - 2, x_{k+1} + 1)$.

(a) Prove that repeated application of such moves to the sequence $(0^m, 2m, 0^m)$ always leads to the sequence $(1^m, 0, 1^m)$ after exactly $\frac{1}{3}(m+1)(m+\frac{1}{2})m$ moves. Here 0^m and 1^m stand for sequences of m 0's and m 1's, respectively.

(b) Prove that, for sufficiently large m, the starting sequence $(0^m, a_1, \dots, a_n, 0^m)$ leads inexorably to the sequence $(0^{m+p}, 1^q, 0, 1^r, 0^{m+n-p-q-r-1})$ for some p, q, and r, if $a_1, \dots, a_n$ are positive integers. Furthermore, p, q, and r can be expressed in terms of $\sum_{j=1}^n a_j$ and $\sum_{j=1}^n ja_j$. How many moves does this transformation require?

E 3268. *Proposed by Dorothy Maharam and A. H. Stone, University of Rochester.*

Does there exist a sequence $\xi_1, \xi_2, \dots$ of positive real numbers such that, whenever $0 < \alpha < \beta$, we have

$$\lim_{k\to\infty} \frac{\#\{n : 1 \le n \le k, \xi_n > \beta\}}{\#\{n : 1 \le n \le k, \xi_n > \alpha\}} = 0?$$

SOLUTIONS OF ELEMENTARY PROBLEMS

Characterizing the Exponential Function

E 3127 [1986, 60]. *Proposed by David Shelupsky, The City College of New York.*

(a) Show that on any real interval $[a, b]$, the exponential function e^x is characterized, up to multiplication by an arbitrary positive constant, by the inequality

$$f(x) < (f(y) - f(x))/(y - x) < f(y), \tag{*}$$

where x and y are real, $a \le x < y \le b$.

(b)* Are the exponentials the only functions on an interval I that satisfy the inequality

$$\min\{f(x), f(y)\} \le (f(y) - f(x))/(y - x) \le \max\{f(x), f(y)\} \tag{**}$$

for $x \ne y$ in I?

Solution to (b) by Bjorn Poonen (student), Harvard University.* The answer is "yes." We prove (b)* by showing that (**) implies the continuity and differentiability of f and by then solving the resulting differential equation. First consider any subinterval $[c, d] = J \subseteq I$, $|J| \le 1/2$, and fix $t \in J$. Then for $x \in J$, (**) implies that

$$|f(x) - f(t)| \le \tfrac{1}{2}|f(x)| \text{ or } |f(x) - f(t)| \le \tfrac{1}{2}|f(t)|,$$

and so $|f(x)| \le 2|f(t)|$ in any event. Thus, for all $x, y \in J$,

$$|f(x) - f(y)| \le 2|f(t)|\,|y - x| = M|y - x|,$$

so f is continuous. But then (**) implies that f is differentiable on (c, d) and $f'(x) = f(x)$, so that $f(x) = ke^x$ for appropriate k and $x \in (c, d)$. If I has length $> 1/2$, write $I = J_1 \cup \cdots \cup J_r$, where $|J_i| \le 1/2$ and $|J_i \cap J_{i+1}| > 0$. Then

$f(x) = k_i e^x$ for $x \in J_i$, and by considering $x \in J_i \cap J_{i+1}$, we see that $k_i \equiv k$ for all i.

As a check, $(ke^y - ke^x)/(y - x) = ke^z$ for some z between x and y, by the Mean Value Theorem, and so $f(x) = ke^x$ satisfies (**).

If f satisfies (*), then it satisfies (**), so all solutions to (*) have the form $f(x) = ke^x$; (*) is satisfied if and only if $k > 0$, and this solves (a).

There were twenty-six correct solutions to part (b)* and eight solutions to part (a) only. Three incorrect solutions were received.

Constructing a Triangle

E 3134, [1986, 132]. *Proposed by Jordi Dou, Barcelona, Spain.*

Provide a Euclidean construction of a triangle ABC, given the median m_A, the bisector t_A, and the angle A.

Solution I by G. Velissarios, Athens, Greece. Consider the triangle ABC to be constructed, as in FIG. 1. Let AM, AD be the median and bisector from A, respectively, with α the measure of the angle. Extend the side AB beyond A to the point R such that $AB = AR$, and let AS be the bisector of the angle at A in the triangle ACR.

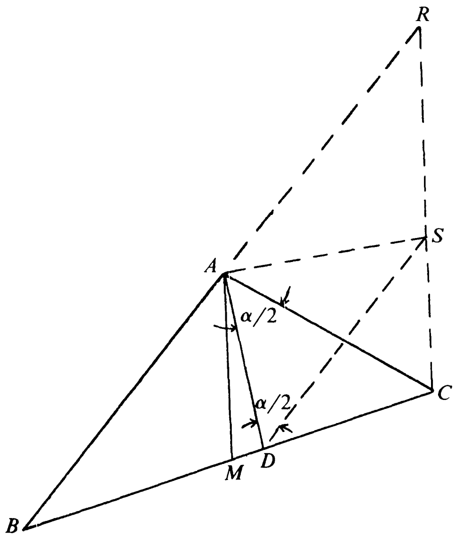


FIG. 1.

We begin by constructing the right triangle DAS. Since AD, AS are angle bisectors at A, we have

$$\frac{DC}{DB} = \frac{AC}{AB} = \frac{AC}{AR} = \frac{SC}{SR}.$$

Therefore, DS is parallel to BR. This tells us that the angle at D in DAS is $\alpha/2$, which enables us to construct DAS. Let $s = AS$.

applying the law of sines to BMA and CMA yields

$$\frac{BC}{2m_A}\left[\frac{1}{\sin(\alpha/2-\delta)}+\frac{1}{\sin(\alpha/2+\delta)}\right]=\frac{1}{\sin B}+\frac{1}{\sin C}.$$

Applying the law of sines to BDA and CDA yields

$$\frac{1}{\sin B}+\frac{1}{\sin C}=\frac{1}{t_A}\cdot\frac{BD+DC}{\sin(\alpha/2)}=\frac{BC}{t_A}\cdot\frac{1}{\sin(\alpha/2)}.$$

Combining these and expanding $\sin(\alpha/2\pm\delta)$ yields an equation that simplifies to

$$\frac{t_A}{2m_A}[2\sin(\alpha/2)\cos\delta]\sin(\alpha/2)=\sin^2(\alpha/2)\cos^2\delta-\cos^2(\alpha/2)\sin^2\delta.$$

Using $\sin^2\delta=1-\cos^2\delta$, we obtain a quadratic equation for $\cos\delta$. The solution to this is $\cos\delta=g+\sqrt{g^2+\cos^2(\alpha/2)}$, where $g=(t_A/2m_A)\sin^2(\alpha/2)$. The solution δ can now be constructed by standard methods.

Editorial Comment. This problem has appeared before as Monthly Problem E1375 [1959, 513; 1960, 185], as noted by L. D. Goldstone and M. Vowe. It also appeared in *Mathematics Magazine* as problem 1054(b) [51(1978) 305; 53(1980) 52–53], as pointed out by M. Vowe and A. Jayakrishna. I. Paasche commented that the problem is solved in a book by K. Herterich, *Dreieckskonstruktionen* (Ernst Klett Verlag, Stuttgart, 1966), 157–158, problem No. 177. I. A. Sakmar provided two solutions, noting that one of them appears in *Exercices de Trigonométrie*, by Frère Gabriel Marie, p. 330. The solutions given above are generally distinct from these.

Also solved by A. Bondesen (Denmark), E. Braune (Austria), L. D. Goldstone, H. Guggenheimer, P. L. Hon (Hong Kong), A. Jayakrishna, O. P. Lossers (Netherlands), L. Kuipers (Switzerland), E. Morgantini (Italy), I. Paasche, I. A. Sakmar (Canada). I. A. Sakmar and O. Yumia (Turkey), S. Selvaraj (student), B. A. Troesch, M. Vowe (Switzerland), and the proposer.

Coverings by a Convex Set

E 3149 [1986, 400]. *Proposed by Louis Funar, University of Craiova, Craiova, Romania.*

Let K be a plane convex set with area a, perimeter p, and diameter d, and let $\lambda=a/(a+pd+\pi d^2)$.

(a) Prove that any finite family of congruent copies of K that covers area A must have a subfamily with pairwise disjoint interiors that covers area at least λA.

(b)* Is this also true with $\lambda=a/(\pi d^2)$?

Solution to (a) by William A. Newcomb, Lawrence Livermore Laboratory, Livermore, CA. If K' is any congruent copy of K, let $E(K')$ denote the set of all points within distance d from K'. The area of $E(K')$ is $a/\lambda=a+pd+\pi d^2$. (Cf. Theorem 14.12 and Exercise 19.9 of Russell V. Benson, *Euclidean Geometry and*

Convexity, McGraw-Hill, 1966.) If F is a finite family of congruent copies of K which covers A, let $\{K_1, \ldots, K_n\}$ be a maximal subfamily of F with pairwise disjoint interiors. Every member of F intersects some K_i and is, therefore, completely contained in $E(K_i)$. Therefore, the area A covered by F is at most n times the area of $E(K)$, so the subfamily covers area $na \geq \lambda A$.

Partial solution of (b) by David B. Secrest (student), University of Illinois, Urbana. Let S be a finite family of congruent copies of K whose union has area A. In the special case in which every pair of sets in S intersect, the diameter of $\bigcup_{B \in S} B$ is less than $2d$. Now the area of any figure with diameter at most $2d$ is at most πd^2 (a small generalization of problem A5 of the 1967 W. L. Putnam Math. Competition). Thus, if $\lambda = a/(\pi d^2)$, we have $\lambda A \leq a$.

Editorial comment. Note that in fact the members of the family need not be congruent in part (a). The argument works as well if all members of the family are convex and have area a, perimeter p, and diameter d, with no further restriction.

Part (a) also solved by the proposer.

ADVANCED PROBLEMS

For instructions about submitting solutions of Problems, which should be mailed before September 30, 1988, see the inside front cover. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label if an acknowledgment is desired.

6572*. *Proposed by David L. Book, Naval Research Laboratory, Washington, DC.*

Show that

$$\int_0^\infty \{1 - e^{-q(t)}\} \frac{dt}{t^{3/2}} = \pi,$$

where

$$q(t) = \frac{1}{\pi} \int_0^\infty \ln(1 + st) \frac{ds}{1 + s^2}.$$

This result appeared in an article in *Zhurnal Eksperimental'noi i Teoretecheskoi Fiziki*, 90(1986) 330 by S. E. Esipov and I. B. Levinson, who stated that they had obtained it numerically and verified it to fourteen significant figures. (At the request of the editors, the Symbolic Computation Group of the University of Waterloo has verified the result to over forty significant figures.)

6573. *Proposed by Gérard Letac, Université Paul Sabatier, Toulouse, France.*

Suppose X is a real random variable which is normally distributed with mean 0 and variance 1. Suppose f and g are functions from R to R such that, for each b in R, the random variable $F(X + b) - g(b)$ is also normally distributed with mean 0 and variance 1. Prove that there exists ε in $\{-1, 1\}$ and c in R such that $f(x) = \varepsilon x + c$ almost everywhere.

SOLUTIONS OF ADVANCED PROBLEMS

A Characterization of Holomorphic Functions

6525 [1968, 574]. *Proposed by Andrew Lenard, Indiana University.*

Let S be an open subset of the complex plane. Prove that a function f is holomorphic on S if and only if

$$\sup_D{}^* \frac{\left|\sum_{j=1}^{n} (z_{j+1} - z_j) f(z_j)\right|}{\sum_{j=1}^{n} |z_{j+1} - z_j|^2} < \infty$$

for every closed disc $D \subset S$. Here the star indicates that the supremum is taken over all finite sequences $z_1, z_2, \ldots, z_n, z_{n+1}$ of points of D (n arbitrary) such that $z_{n+1} = z_1$ and at least two of the z_j are distinct.

Solution by Thomas L. McCoy, Michigan State University, East Lansing, MI. First let f be holomorphic in the closure of the disc D, and M an upper bound for its derivative there. Let

$$C = \bigcup_{j=1}^{n} L_j,$$

where L_j is the line segment from z_j to z_{j+1}. By convexity, $L_j \subset D$. We have then

$$\begin{aligned}\sum_j (z_{j+1} - z_j) f(z_j) &= \int_C f(z)\,dz - \sum_j \int_{L_j} \left(f(z) - f(z_j)\right) dz\\ &= -\sum_j \int_{L_j} \left(f(z) - f(z_j)\right) dz,\end{aligned}$$

whence

$$\begin{aligned}&\left|\sum_j (z_{j+1} - z_j) f(z_j)\right|\\ &\quad\le \sum_j |z_{j+1} - z_j| \sup_{z \in L_j} |f(z) - f(z_j)|\\ &\quad= \sum_j |z_{j+1} - z_j| \sup_{z \in L_j} \left|\int_{z_j}^{z} f'(w)\,dw\right|\\ &\quad\le \sum_j |z_{j+1} - z_j| M |z_{j+1} - z_j|.\end{aligned}$$

Thus holomorphy of f in S does indeed imply an inequality of the desired kind for any closed disc lying in S.

Conversely, assume

$$\left|\sum_{j=1}^{n} (z_{j+1} - z_j) f(z_j)\right| < M \sum_{j=1}^{n} |z_{j+1} - z_j|^2$$

is valid for the closed disc D. Then $n = 2$ yields

$$|f(z_2) - f(z_1)| < 2M|z_2 - z_1|,$$

so f is continuous in D.

Now let Γ be any simple closed rectifiable curve lying in D. Then the integral of f over Γ is the limit of sums $\Sigma(z_{j+1} - z_j)f(z_j)$ as $\sup|z_{j+1} - z_j| \to 0$, the z_j traversing Γ. Let $s(\Gamma)$ be the length of Γ. Choose the z_j on Γ so that $|z_{j+1} - z_j| < \delta$ for all j, and $\Sigma|z_{j+1} - z_j| < s(\Gamma)$. Then

$$\begin{aligned}\left|\sum f(z_j)(z_{j+1} - z_j)\right| &< M\sum |z_{j+1} - z_j|^2\\ &< M\delta\sum |z_{j+1} - z_j|\\ &< M\delta \cdot s(\Gamma),\end{aligned}$$

and since δ is arbitrarily small we conclude that

$$\int_\Gamma f\,dz = 0.$$

By Morera's Theorem f is holomorphic in D. Since D is an arbitrary disc contained in S, f is holomorphic in S.

Remarks by the proposer. If the supremum of the problem is taken only over sequences of length $n = 2$, one obtains the class of Lipschitz-continuous functions on D, as mentioned in the course of the above solution. If n is unrestricted, one obtains the class of holomorphic functions. But what if we take the supremum only over sequences of length $n = 2, 3, \ldots, N$ for some fixed N? This defines a certain class C_N of functions on D, so that C_2 is the set of Lipschitz-continuous functions and

$$\bigcap_{N=2}^{\infty} C_N$$

is the set of holomorphic functions. Can one determine in some effective sense what C_N is?

Editorial Comment. All solutions received were essentially the same as McCoy's.

Solved also by O. P. Lossers (The Netherlands), Chr. A. Meyer (Switzerland), Victor Manjarrez, and the proposer.

REVIEWS

EDITED BY JOSEPH KONHAUSER

Department of Mathematics, Macalester College, St. Paul, MN 55105

Mathematics as Problem Solving. By Alexander Soifer. Center for Excellence in Mathematical Education, Colorado Springs, CO, 1987. viii + 117 pp.

CECIL ROUSSEAU
Department of Mathematical Sciences, Memphis State University, Memphis, TN 38152

In the terminology of Paul Erdös, **The Book** is that transfinite collection of all theorems and their best possible proofs. Perhaps the essence of **The Book** is to be found in its section entitled *Elementary Problems and Solutions* (or words to that effect). Here are those problems which require no elaborate formal background but whose solutions provide glimpses of mathematics at its finest. Here, to be sure, is L. M. Kelly's solution of the problem posed by J. J. Sylvester and first solved by T. Gallai.

Prove that if n points in the plane are not on one straight line, then there exists a straight line containing exactly two of the points.

Here are examples from George Pólya and other masters of the art of problem solving. Candidates are to be found in the *Elementary Problems* section of the *American Mathematical Monthly* and in the various problem-solving competitions sponsored by the MAA. The best of these are recognized as coming "straight from **The Book.**"

Elementary problem solving at its best plays the distinctive role of uniting generations of mathematicians. Young Olympiad and Putnam caliber students share with "Uncle Paul," the world's most famous $2\frac{1}{2}$-billion-year-old mathematician, common appreciation for problems in elementary number theory and combinatorics. Mathematical fashions change, but **The Book's** geometry problems are timeless. The best problem solvers of a new generation will find in them the same satisfaction as did the best of their predecessors.

What makes a solution suitable for **The Book?** What is it that we find so attractive? Mathematicians often speak of "beauty", "elegance", etc. in reference to theorems and their proofs. Generally speaking, such talk leaves the nonmathematical fellow citizen totally bewildered. There is perhaps something more elemental to a solution from **The Book.** That is the element of surprise. Our sense of elegance may be so pronounced because we are caught off guard by the proof's crucial idea and are thus surprised by its "elegant" twist. In this respect, there is much to connect mathematics and humor. A good solution is like a good joke; the right touch produces a classic. The quintessential introduction of a new problem or proof to a colleague may be "Have you heard the one about... ?"

An illustration is in order. The following *sleeping mathematicians problem* was on the 1986 USA Mathematical Olympiad.

During a certain lecture, each of five mathematicians fell asleep exactly twice. For each pair of these mathematicians, there was some moment when both were sleeping simultaneously. Prove that, at some moment, some three of them were sleeping simultaneously.

Members of the committee which constructed the Olympiad had in mind a "nice" solution using interval graphs. However, several of the students taking the exam were more insightful than the professors. Their solution goes like this.

> Suppose that no three mathematicians were ever asleep at the same time. Then there are $\binom{5}{2} = 10$ *disjoint* time intervals in which two mathematicians are simultaneously asleep. These intervals are each initiated by the event of a mathematician falling asleep and there are exactly $5 \times 2 = 10$ such events. Hence, there are 10 *distinct* times at which a mathematician falls asleep, each such time initiating a different period of common dozing. This is absurd; just consider the *first* time a mathematician falls asleep.

No fancy formulas, no interval graphs, just a surprising and witty solution—and, of course, a good joke on the professors.

Earlier, we mentioned L. M. Kelly's solution of Sylvester's problem. The proof is like a classic story. It may have been told many times in the past, but deserves to be retold.

> Out of all such choices, choose a point P_i and a line L not containing P_i but containing at least two of the remaining $n - 1$ points so that the distance from point to line is minimized. (Justification: not all of the points are on the same straight line and the sets involved are finite.) Let Q denote the point on L which is nearest P_i. If L contained three or more of the n given points, then two of these points would lie in one of the two rays obtained by dividing L at Q. Suppose that this is the case and let P_j and P_k denote two such points, with P_j being the nearer of the two to Q. Since the distance from P_j to the line through P_i and P_k would then be less than the distance from P_i to L, this is impossible. Thus, L cannot contain three or more of the points. It contains precisely two.

Retelling the best solutions and sharing the secrets of discovery are part of the process of teaching problem solving. Ideally, this process is characterized by mathematical skill, good taste and wit. It is a characteristically personal process and the best such teachers have surely left their personal marks on students and readers. Alexander Soifer is a teacher of problem solving and his book, *Mathematics as Problem Solving*, is designed to introduce problem solving to the next generation. By external examination, it is a fairly modest book (117 pp) covering standard elementary topics and containing about 200 problems. However, it is somewhat more than this. It is the reflection of the author's personal experiences and it is an expression of his desire to share what he has learned from **The Book**.

Toward a Lean and Lively Calculus: Report of the Conference / Workshop to Develop Curriculum and Teaching Methods for Calculus at the College Level. Donald G. Douglas, Editor. The Mathematical Association of America, 1987. xxi + 249 pp.

RICHARD W. HAMMING
Department of Computer Science, Naval Postgraduate School, Monterey, California 93943

There has recently been a lot of noise about revising the calculus course to meet modern needs, and it is natural that the needs of society change. But it is also all too common that the changes are not recognized by those in charge until many years later. For example, E. E. David once remarked to me that our society has passed from a manufacturing society to a service oriented one, but the appropriations for R & D, especially those from Washington, show no recognition of this fact. I similarly observe that we have moved from a society dominated by mechanics and electrical circuits in engineering to one dominated by probability and statistics, but that the mathematical curriculum has not changed accordingly. And, I would say, for about the same reasons.

The students have also changed. One change that has often been noted is that as we have broadened the base of enrollment we have lowered the average level of competence. The recent burst of the "New Mathematics" has not helped in their earlier education as both they and their teachers remain, to a great extent, demoralized. But the students have changed in other ways—they recognize that they do not live in a world of mechanical things (as I did in my youth) but rather a very different world to which almost all that they read in the current calculus books appears to be irrelevant. The artificial problems that appeal to us generally do not appeal to them with their sense of "relevance." They remain passively engaged in the courses we currently teach. There is little identification of the content with their possible lives, but rather mathematics is a chore, a hurdle to be got by as gracefully as possible. We must find applications that appeal to the students as they are at the time they are in class. I expect that the economics problems I see in the calculus books seem to them (as they do to me) little more than empty stuff, and the biological examples are not much better! The problems must have some real connection with life as *they are leading it*, involve things that they have heard about and are interested in.

There is another fundamental difficulty with the teaching of mathematics as it is presently done. Any systems engineer knows that if you optimize the parts of a system then almost surely the system performance will dègrade. We have finally managed to hone the individual courses like calculus and linear algebra so that they are optimum for themselves—and in the process the teaching of mathematics has degraded. Until each course in mathematics is designed to support the whole system we will have this counterproductive result. I occasionally teach, for the mathematics department, a classic (static) linear algebra course while for the electrical engineering department I teach dynamic linear algebra under the name "Digital Filters." Generally speaking, the static course is completely unaware of the dynamic side of linear mathematics.

We now turn to the basis for complaining about the current calculus courses as taught. There are two sources possible: those from within and those from without the system. Taking first those from within, we see that there is in fact just one

standard calculus book with but very minor differences (including the proofs given!) in wide use, while excellent books written by first class mathematicians that deviate very far are simply ignored. Since it is the professors who choose the texts we have to assume, no matter what they claim, that the books being used represent what they want. With the power of choice rests the responsibility. There are other books around, but the professors will not choose them! From this you conclude that there is, in fact, no large scale complaint from within.

We are a democracy, and this, along with academic freedom, means that any college that wants to try new things has the power to do so in spite of screams of not being compatible with the rest of the system. Reform can be accomplished in this country by the democratic process of the individuals voting as they believe and not by central dictatorship (as was once tried by a few pure mathematicians with great financial backing to force the New Mathematics on the world). Until the professors (who have the power to select) change, there will be little change no matter how much noise is made or how many committees and reports are produced.

The complaints from the outside have been mainly from the computer science departments who fancy that they must have a course in discrete mathematics—but there is little agreement on what is wanted. Most of the books so far produced look, at least to this reviewer, like a hodgepodge of isolated topics that are often only loosely connected. Indeed, often a notation that is introduced early in the book is simply abandoned later!

Some of the loudest advocates in the past for abandoning calculus for a discrete mathematics course have gradually toned down their complaints as they have paused to listen to the simple fact that discrete mathematics carried very far usually leads to generating functions (which imply series, integration and differentiation), and the generation of a new identity from an old one uses the method of calculus (or else a completely contrived derivation that is meaningless to the student). These complainers seem now to be more in agreement that what they want is the calculus with more discrete mathematics incorporated, like generating functions and difference equations. But, Apostol's magnificent books have long had a lot of discrete mathematics in them, so the texts are there (including one by this reviewer) but are seldom used.

I doubt that the computer science departments will get the mathematics departments to change much. The American Mathematical Society forced the formation of the Association for Computing Machinery long ago by not even allowing a single session on numerical methods. The computer science departments may very well take up teaching discrete math courses, but they also very likely will not be satisfied with what they do in the long run, unless they include a lot of the continuous calculus.

Any competent mathematician knows that the use of complex variables in number theory has been very fruitful—the extremes of discrete and continuous (analytic in fact) meet here successfully to the advantage of both. Indeed, there is a whole field known as analytic number theory that blends the two. Similarly, as noted above, combinatorics in any depth rests on generating functions. It appears to this reviewer that any attempt to enforce a strict separation is damaging to both.

We have yet to discuss what the calculus is. To pure mathematicians it is an interesting exercise in the real line and mappings. To scientists and engineers it is a powerful tool kit of methods (why else the name?) to be applied in many situations.

But here we run into trouble. The slant given by the current mathematical texts does not fit what is wanted and in fact leads to ridiculous results. For example, the popular (Bourbaki?) definition of a function as a set of ordered pairs runs right into the simple desire to count multiple zeros as multiple zeros (as needed, for example, in linear differential equations with constant coefficients). The static definition is also inappropriate to the calculus which is the study of change, (dynamics). A function in the calculus is more like a curve being traced by a moving point than it is a set of pairs of numbers. Newton used the word "fluxions" to describe his dynamic picture of what is going on. From Newton through Euler, it was the dynamic view of function that gave them inspiration. The static definitions may have provided some rigor, but at the cost of a meaningful treatment as far as the engineers and scientists are concerned, let alone the students! The current books are a poorly digested mixture of the static New Math and the dynamic calculus of change—and you see it in most of the books when the author forgets what he purports to believe and lapses into the dynamic view.

But there is another aspect of the calculus that is sometimes recognized. The expression "mathematical maturity," whatever it means to you, is probably achieved first, if at all, in the calculus course where there is so much generality that the student cannot get by with blind memorization, but must come to terms with the manipulation of symbols obeying given relationships—one important aspect of mathematics! For example,

$$\int_a^b f(x)\,dx = \int_a^b f(t)\,dt$$

and the variable x or t does not matter. Any extensive simplification of the contents of the calculus course is in grave danger of losing this essential step in the mathematical development of the student. We can change the contents, to be sure, but we need to be very wary of making it "easy."

This aspect of the calculus can be stated in another way. I once went to my favorite dean who was very concerned with teaching and told him the following story. I was, so I said, teaching a weight-lifting class, and the final test was to lift 250 pounds. I saw that many students had to repeat the course, that many got discouraged and dropped out, and that few passed. One night, worrying about this, I decided that I should cut the weights in half, and the final test would then be to lift one set of 125 pounds and then the second set, thus in the end lifting the 250 pounds. I went on to say, "When I make a presentation in class of some material significantly easier, am I not cutting the weight in half?" We are trying in the long run to develop the "mental muscles" of the students, and it is not the particular content and its presentation that is the long-range goal. Thus, making the course easier and easier for the student to learn is perhaps counterproductive to the development of the student's mind. This is one of the many reasons why I doubt the wisdom of trying to teach the calculus to the brighter high school students—I fear that the weights are indeed lifted, but that there is deception in the process. There is not the development of the mind which is the goal. I am well aware that it is difficult to measure this elusive development I am speaking of, but to turn around and measure irrelevant things and base one's judgment on them seems to me to be sheer folly—though that is what is done in most educational circles!

It is not easy to decide if the textbooks are thicker because of more new and essential material or because of more worked examples, more pictures, more

problems, more introductions and summaries, and more background reviews of needed mathematics. I recall that my college text had the topics of envelopes, evolutes, etc. Envelopes are essential to the understanding of the solution of first order nonlinear ordinary differential equations because it is the envelope, which is not in the so-called "general solution," that often is wanted in practice. The idea is thus a fundamental one, but has been dropped from the texts. I cannot do a simple objective survey of the books by counting lines devoted to the various aspects because this method founders on the simple fact that an author may use what appears to be a worked example and follow with a few lines of text saying "From the example follows the general case of... ."

This reviewer, based on his 30 years as an industrial mathematician, tends to feel strongly that the ability to juggle symbols as the pure mathematician does without regard to the immediate meaning of the symbols is but half of being a mathematician. The other half is the ability to apply the mathematics to the real world. The students also seem to feel that playing the mathematical game (as many mathematicians will claim they are doing when pushed to the wall about the meaning of mathematics) is no more meaningful than playing bridge and chess as a lifetime career. The calculus is one of the great fields of mathematics where applications are easy to find that can be meaningful to the students as they are.

I need to discuss another well known aspect of mathematics—that it is the art of abstraction, generalization, and extension. It seems to me that this is not best taught by starting with abstract postulates, but rather by beginning with the concrete and teaching the process of abstraction. Thus in class just a few days ago we had the problem of finding the maximum area of a rectangle that can be put inside a certain ellipse. I got the answer to it, and of course it was merely the answer. Then I turned to the general ellipse, found the solution for it and showed that: (1) it matched one's ideas for a circle, (2) it gave the earlier specific case, (3) it was much more understandable than the specific case, and (4) it had all the symmetry and dimensional analysis it should have. They could then see that it was simply the case of a square in the circle tilted until it looked like the ellipse they originally had been given. The abstraction shed light on the specific case that was obscured by its particular details. We need to do such abstractions regularly if we are to teach mathematics and not just cover the ground.

There is yet one final aspect of what calculus is and is not. It is in many respects the systematic evasion of hard computation: how to go right to the answer without fumbling around! Often the calculus gives you the method. Thus trying to blend numerical analysis and the calculus is a lot like mixing water and oil. A small amount of computing is useful, though trying to claim that computing an infinite series to show its convergence is plain dangerous—the harmonic series will converge (slowly to be sure) on a computer to a finite value, but in mathematics the limit is infinite. Furthermore, any serious numerical computation uses floating point numbers (von Neumann to the contrary!) and most mathematicians are very ill prepared to understand them. Floating point numbers are tricky, very unlike the standard number system of mathematics, and the teachers who do not master floating point arithmetic will simply look like fools to knowledgeable students who have done extensive computation on their computers.

Let me now turn to the book being reviewed, *Toward a Lean and Lively Calculus*, which is the result of a Symposium held January 2–6, 1986, at Tulane University

with the 25 distinguished attendees, listed in the front. A statistician friend of mine used to say, "Let me pick the sample, and I will let you do the analysis." So I looked closely at the names and where they were from. There were 16 from mathematics departments, one I recognized as an engineer, a lot of professional educators at various levels, a professor from Courant Institute, and one high school teacher—but none from industry that I could recognize. Well, given the sample, it is not hard to deduce, without reading much, what the book will contain. The reports and papers are good as far as they go. For example, Stein is clearly right when he tries to find out where the complaints are coming from, and finds none in the standard places, and Renz agrees that the discrete is being oversold.

Being in a rather protected place in the educational system, I had not known that classes of 500 or more calculus students are widespread. I was appalled at first, but when I thought it over, there is no absolute reason against classes of this size, provided they are well done. But the fact that graduate students manage most of the quiz sections, and probably not the best mathematical and teaching professors are used to give the lectures, makes one worry. One proposal in the report that the graduate students be given courses in teaching will have no effect unless the graduate students perceive that it will count toward getting their Ph.D. degrees—and I cannot believe that mathematics departments will do this in the near future. What is seldom thought about is this; only one faculty person sees the class size of 500 or more, but 500 students see that size, which is quite a different matter! I have had some time to think the idea over and to talk to friends about it, and I cannot believe that there are many places (there may be a few) where it is well done. I have a sick feeling when I consider how such courses probably work out in practice.

If you think that the calculus is merely technique and the course is a drill place, then maybe you can defend that size, but if you think that it is an educational experience then you will have to reconsider matters. Of course it can be defended on the grounds that it makes the students learn for themselves and that this is good, but I still have a very heavy heart when I think that this is what we are doing to eager, expectant beginners in the field of mathematics. No wonder we have so few math majors!

The failure rate of 50% in many places also makes me wonder. To the outsider it appears to be a heartless method without any attempt to weed out the losers before starting. One would think that a good precalculus course, or else a stiff qualifying exam in order to be excused, would be widespread, and once a student was in a calculus course there would be a reasonable probability of passing. But according to the report, failure rates of 50% seem to be regular—and there seems to be little apology for it. It is foolish for the mathematics departments to accept the onus of eliminating the less able students from the university!

I suspect that the complaints about the calculus course arise from these monster classes and the way they are handled—and from the professors who do not want to teach the calculus course, though I myself clearly believe the calculus course should be changed.

As I said, there is much wisdom in the report; Renz for example seems to recognize that we are at a local optimum of teaching the calculus, and that small reforms will not be effective—but I doubt that the installed professors are willing to go for a big change; it is too much trouble!

Lax is enthusiastic about the importance of the continuous mathematics, and he is right; still there is an increasing amount of discrete mathematics, probability, and statistics in science and engineering, and it will not go away, let alone go away from the coming fields that will increasingly need a good, solid knowledge of mathematics. One cannot, in my opinion, do serious statistics using only the discrete, since the "statistics of a statistic" is a fundamental idea that is essential to the understanding of statistics, and this involves the continuous!

Maurer is also very good; but it becomes invidious to name some and not others, while a complete listing of all the names seems foolish. Even the preface is full of wisdom. Yet nowhere did I see the recognition that a calculus course is only part of a system of mathematical education and that the optimization of it is counterproductive to the mathematical education of society since it is well known in the field of systems engineering that optimizing a component generally degrades the system's performance. Calculus is needed both broadly, and in much greater depth, than was necessary in the past if people are to participate in the evolution of our society at other than a superficial level. We indeed live in an increasingly technical society in all fields, from science and engineering through the hard sciences and into the softer sciences, hence mathematics, along with probability and statistics, is now entering, inevitably, into even the humanities and arts. Mathematics is the language of clear thinking, and where clarity is wanted there is the need for mathematics. Calculus, being the basic language for describing change, will remain the basic intellectual tool for understanding our increasingly complex, changing society. Reform of the calculus course is necessary but I believe it must come from within the mathematical community and cannot be imposed from without. Will the mathematicians respond? I want to believe that they will, but I doubt it!

Decompositions of Manifolds. By Robert J. Daverman. Academic Press, 1986. xi + 317 pp.

JAMES W. CANNON
Department of Mathematics, Brigham Young University, Provo, Utah 84602

Applications

At least three famous problems in geometric topology have been solved in recent years by arguments involving decompositions of manifolds. These problems are the double-suspension problem for homology spheres, the characterization problem for topological manifolds, and the 4-dimensional Poincaré conjecture. Daverman's book includes a proof of the first of these and gives a beautiful exposition of the ideas, the central examples, and the types of arguments involved in the decomposition theory related to all of these problems. As a highlight, Daverman proves the cell-like approximation theorem of Edwards which has been known for many years now but which has never before appeared in print.

Elementary Examples

Consider a dozen eggs. Their union might be considered an infinite topological space X with twelve components. But it is equally natural to consider each of the

twelve eggs as a single point which together form a space Y having exactly twelve points and the discrete topology. The space Y is formed from the space X by *identification* or *decomposition*.

The general construction is this. Let X be any space, and let $\sim$ be any equivalence relation on X. Let $Y = X/\sim$ denote a new space whose points are equivalence classes of points of X. Declare a subset U of Y to be open in Y if the union of the equivalence classes of U form an open subset of X. Then Y is called a *decomposition space* of X. Decomposition theory compares the properties of X and Y. There is, of course, a natural continuous surjection π: $X \to Y$ which takes each point of X to the equivalence class which contains it.

The projection map π: $X \to Y$ of the preceding paragraph suggests a standard way of obtaining a decomposition space. Let X be any topological space, Y a set, and f: $X \to Y$ a surjection. Declare $x, y \in X$ to be equivalent if and only if $f(x) = f(y)$. We may then consider Y as the set of equivalence classes and endow Y with the identification topology. Then the map f becomes the projection map.

As an example of a decomposition consider the case where X is a plane and the only equivalence class having more than one point is a (closed) arc in X. Then Y is also a plane, as the reader may prove for himself. On the other hand, if the only equivalence class having more than one point is a simple closed curve in X, then Y is homeomorphic with a plane to which a tangent bubble has been attached.

Daverman's book considers the case where X is a topological manifold and $\sim$ has especially nice properties. There are two typical hypotheses on $\sim$. The first is that $\sim$ should be *upper semicontinuous* (usc). One way of stating this is that Y should be Hausdorff. Another is that the image of a closed subset of X should be closed in Y. The second typical hypothesis on $\sim$ is that it should be *cell-like*. This means that each equivalence class should be a compact set having the Čech-homotopy properties of a point; that is, this compact set should be contractible to a point in each neighborhood of itself. These two hypotheses taken together ensure that X and Y are hereditarily of the same homotopy type; that is, if U is an open subset of Y, then the inverse image of U in X is homotopy equivalent with U.

The theory is full of beautiful and surprising results.

Sources of Decomposition Theory

Here are three of the beautiful results that led to the development of the theory of decompositions of manifolds.

R. L. Moore's theorem. If f: $S^2 \to X$ is a surjection from the 2-sphere S^2 onto a Hausdorff space X such that, for each $x \in X$, $S^2 \setminus f^{-1}(x)$ is nonempty and connected, then X is also a 2-sphere. Moore's students asked what the appropriate generalization would be in higher dimensions.

R. L. Wilder's study of manifolds. J. W. Alexander defined a topological 2-sphere S in Euclidean 3-space E^3 in such a way that the union B of S and its interior Int S in E^3 was not a 3-ball. We call the set B an *Alexander horned cube*. R. L. Wilder noted that, if one were to sew two copies of B together along their 2-sphere boundaries by a homeomorphism of those boundaries, then the resulting union had all of the standard algebraic properties of the 3-sphere, just as if one had sewn together two copies of the 3-ball. Wilder asked whether the sewing was in fact the 3-sphere.

P. A. Smith's theory of periodic mappings. P. A. Smith proved that if one were to consider a homeomorphism $T: S^3 \to S^3$ from the 3-sphere S^3 to itself whose square $T \circ T$ was the identity, then the fixed point set of T, that is, the set of points x such that $T(x) = x$, was also a topological sphere, possibly of lower dimension. Smith asked whether every such periodic map was conjugate to an orthogonal map.

Notice that none of these three results mentions decompositions of manifolds. But each of them suggested an important problem or field of study that led to the study of decomposition spaces. This transition to decomposition spaces occurred in large measure because of the work of R. H. Bing.

Tribute to R. H. Bing

It was R. H. Bing who outlined all of the standard problems, examples, and techniques in the study of decompositions of manifolds.

Pointlike decompositions. R. L. Moore's theorem can easily be restated as a decomposition space theorem: if X is a cell-like usc decomposition of the 2-sphere, then X is a 2-sphere. The students of R. L. Moore asked what the appropriate generalization of this remarkable theorem would be in higher dimensions. G. T. Whyburn suggested a condition on the point preimages of a map and their embedding in the domain of the map in order that the domain and range be homeomorphic. He called a set *pointlike* in a space if its complement was homeomorphic with the complement of each and every point of the space. Two related conditions are *cellularity* and *cell-likeness*. A set in an n-manifold is *cellular* if it is the intersection of a nested sequence of n-cells. In Euclidean space a set is point-like if and only if it is cellular. A set in an n-manifold is *cell-like* if it contracts in each neighborhood of itself. Every cellular set in a manifold is cell-like. Bing was the first to give an example of a point-like (cellular) decomposition of a Euclidean space such that the decomposition space was not a manifold. This example was named the *dogbone space*.

The Alexander horned sphere, Wilder's problem, Bing, and the rubber bands. Bing was interested in the question implicit in P. A. Smith's theory of periodic maps: is every involution of the 3-sphere conjugate to an orthogonal involution? Bing realized that if Wilder's sewing of two Alexander horned cubes were S^3, then there would be a nonorthogonal involution of S^3 which interchanged the two cubes. Bing then proceeded to realize Wilder's sewing as a point-like decomposition space of the 3-sphere S^3. In turn, Bing hit upon an idea for showing that the decomposition space was in fact S^3; the nondegenerate elements of the decomposition were described as infinite intersections of solid tori; he realized that it would suffice to arrange the solid tori in such a manner that they were near the union of their original positions yet had small diameter. He took rubber bands as models of solid tori, arranged these rubber bands in the desired configuration, and then proceeded to play with the rubber bands in order to see whether they could be "shrunk." His clever solution to the shrinking problem added a multitude of new ideas to the theory of manifolds. In the process, Bing answered the Wilder question in the affirmative and demonstrated the existence of nonorthogonal involutions of S^3.

The dogbone space and minimal counterexamples. Bing typically worked by induction on the complexity of the examples his techniques would handle. He would

try to prove a theorem for two hours, develop a counterexample for the next two hours, and cycle the procedure. His techniques often resulted not only in the correct theorem but in the best possible example to show the limits of that theorem. "A good theorem," he delighted in saying, "is one that is almost false." In a sense, Bing's dogbone space was a minimal counterexample to the conjecture that cellular decompositions always had manifolds as decomposition spaces.

Manifold factors. Shapiro showed that the contractible 3-manifold defined by Whitehead which was not Euclidean 3-space was nevertheless a factor of Euclidean 4-space. Bing, who had not seen the proof, discovered a proof for himself and generalized his own proof to show that the dogbone space was also a factor of E^4. Bing's procedure was very powerful and eventually led people to the realization that manifold factors and generalized manifolds were essentially the same class of spaces.

Highlights of Decomposition Theory

We here describe three results of considerable interest that were solved by use of the mature theory of decomposition spaces of manifolds.

The double-suspension theorem. A finite complex that has the local and global homological properties of the n-sphere is called a simplicial homology sphere. Poincaré defined the first homology sphere D which was a true topological manifold but which was not a true topological sphere. His homology sphere was of dimension 3 and was clearly not a sphere because it was not even simply connected. If D is embedded in a Euclidean space E^k, if E^k is considered as a coordinate hyperplane of a still higher dimensional Euclidean space E^{n+k}, and if S denotes the standard unit sphere in the orthogonal complement E^n of E^k, then one may form the join $S * D$ of D with S. The join $S * D$ is simply the union of all straight line intervals joining a point of S with a point of D in E^{n+k}. The join $S * D$ is called the nth suspension of D. If D were a true sphere, then $S * D$ would be a true sphere of higher dimension. However, since D is not simply connected, for $n = 1$, $S * D$ cannot possibly be a true sphere, not even a manifold, for it is not like a manifold at the two points of the 0-sphere S. People noted, however, that for $n = 2$ the join $S * D$, which was called the double-suspension of D, had all of the standard local and global invariants of the $k + 2$-sphere. Milnor asked, as one of his famous problems, whether it could possibly be a true topological sphere. It was clear that if it were a sphere, then it would be impossible to define a piecewise linear homeomorphism from the finite complex $S * D$ to the standard sphere. R. D. Edwards and J. W. Cannon finally resolved the problem by decomposition space techniques by showing that the double suspension of any homology sphere is indeed a topological sphere. The homeomorphisms they described did indeed involve infinite processes, indeed shrinking techniques of the type suggested by Bing.

The 4-dimensional Poincaré conjecture. The most famous problem in geometric topology is clearly the Poincaré conjecture: a compact, topological manifold without boundary is a sphere if it has the homotopy type of a sphere. This conjecture was confirmed for manifolds of dimension $\geqslant 5$ during the sixties by Smale, Zeeman, Stallings, and others. Freedman confirmed the conjecture for dimension 4 in the early eighties. Freedman's work depended heavily on decomposition space techniques.

Characterization of manifolds. Examples of decompositions of manifolds, particularly examples due to Bing, showed the existence of spaces that in all of their standard algebraic properties were like manifolds but which were not topological manifolds. The examples made it apparent that there would be no simple characterization of the topological manifold in terms of simple, checkable topological properties short of the local coordinate definition. It came as a great surprise, therefore, when decomposition studies indicated the probability of the following conjecture: a generalized n-manifold having the disjoint disks property, $n \geqslant 5$, is a topological n-manifold. A generalized n-manifold M is a Euclidean neighborhood retract whose local homology coincides at each point with the local homology of Euclidean n-space. The space M has the disjoint disks property if maps $f, g\colon B^2 \to M$ can be approximated by maps $f', g'\colon B^2 \to M$ having disjoint images. Cannon stated the conjecture early in 1977 and proved many special cases. The general case is still partially in question, though major steps presented by R. D. Edwards and F. Quinn which almost complete the proof seem firmly established.

Daverman's Book and the Three Periods of Decomposition Theory

Daverman himself is one of the modern masters of decomposition theory. He divides the history of this subject into three rough areas.

The Moore period. The first period dealt primarily with decompositions of 2-manifolds. It was typified by the theorem of R. L. Moore mentioned above. The main technique centered in an axiomatic characterization of 2-manifolds.

The Bing period. The second period dealt primarily with 3-manifolds. The central technique was the shrinking of decompositions. The principal practitioner was R. H. Bing. The most beautiful results were a catalogue of beautiful examples that were not shrinkable but which were factors of 4-manifolds. Daverman states the following positive theorem as prototypical.

THEOREM (Bing and Armentrout). *Let G be an upper semicontinuous decomposition of a 3-manifold M into cellular sets. Then M/G is homeomorphic to M if and only if G is shrinkable.*

The Edwards period. Daverman associates the third period with the work of Edwards. The central result is Edwards' cell-like approximation theorem.

THEOREM (CELL-LIKE APPROXIMATION). *Let G denote an upper semicontinuous decomposition of an n-manifold M, $n \geqslant 5$, into cell-like sets. Then the decomposition map $\pi\colon M \to M/G$ can be approximated by homeomorphisms if and only if M/G is finite-dimensional and satisfies the following disjoint disks property: any two maps of the 2-cell B^2 to M/G can be approximated by maps having disjoint images.*

The bibliography of this book reads like a listing of my mathematical friends and neighbors during a period of my life when I lived in a different mathematical neighborhood, city, and state. I thank those friends and neighbors for years of friendship and pleasure. I thank Bob Daverman for recording the family history.

TELEGRAPHIC REVIEWS

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T: Textbook	P: Professional Reading	1–4: Semesters
C: Computer Software	L: Undergraduate Library	**: Special Emphasis
S: Supplementary Reading	13: Grade Level	??: Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

General, P, L.** *Encyclopaedia of Mathematics, Volume 1: A-B.* Ed: M. Hazewinkel. D Reidel (US Distr: Kluwer Academic), 1988, ix + 488 pp, $149. [ISBN: 1-55608-000-X] First of ten volumes of a translation with editorial updates of the Soviet *Mathematical Encyclopaedia,* written in the 1970's and originally published during 1977-1985. (Original Russian edition of 150,000 is sold out.) Contains surveys for non-specialists (e.g., analytic number theory, axiomatic method, basis), short articles on specific results or problems (e.g., Balayage method), and definitions, all arranged alphabetically with extensive cross referencing to MR numbers. Most items contain references, frequently supplemented by the editors to include current Western sources. Final volume will be devoted to extensive indices cross-referencing terms and MR classification numbers. LAS

General, S(13). *Inequalities.* P.P. Korovkin. Transl: Sergei Vrubel. Little Math. Lib. MIR (US Distr: Imported Pub), 1986, 72 pp, $1.95 (P). [ISBN: 0-8285-3221-4] Cute little book that explores various inequalities that pop up in mathematics. The format is to introduce a definition or a certain inequality (for example, the Bernoulli inequality), state and prove a fact or two, then follow with some solved problems. This would be a nice book to use in a high school or freshman problem solving class. (*First Edition,* TR, October 1977.) LC

General, S(13). *Dictionary of Mathematics Terms.* Douglas Downing. Barron's, 1987, xiv + 241 pp, $8.95 (P). [ISBN: 0-8120-2641-1] Definitions for over 600 terms from ***abacus*** to ***zero.*** Some cross-referencing. Uneven in accuracy, clarity, completeness, and utility. Symbol list for algebra, calculus, geometry, vectors, sets, and logic. Brief logarithms, trigonometry, and statistics tables. Handy, but far from the last word. A compact 4 3/16" x 7". Good buy for high school libraries and college students. JK

General, P. *Fourteen Papers Translated from the Russian.* V.I. Arnol'd, *et al.* AMS Transl. Ser. 2, V. 137. AMS, 1987, v + 129 pp, $46. [ISBN: 0-8218-3113-5]

Elementary, T(11: 1). *Intermediate Algebra, Second Edition.* Daniel L. Auvil. Addison-Wesley, 1987, xi + 593 pp, $31.95. [ISBN: 0-201-11046-6] Clearly written in a conversational and unitimidating manner. Presents practical uses of the material. Topics include: real numbers, polynomials, inequalities, rational expressions, rational exponents, complex numbers, graphing, systems of linear equations, conic sections, functions, logarithms, sequences, and series. Plenty of problems follow each section. LW

Mathematics Appreciation, T(14-16), S, L**.** *The Problems of Mathematics.* Ian Stewart. Oxford U Pr, 1987, 257 pp, (P). [ISBN: 0-19-289182-0] A lively, non-technical portrait of important themes in mathematics, each traced from ancient roots to yesterday's research: primes ("equations with holes"), geometry, fractals, groups, knots, computability ("the used axiom salesman"), and much more. Witty and creative, Stewart writes with authority in the manner of a well-researched travel guide: he conveys enthusiasm and insight on every page. LAS

Finite Mathematics, T(13: 1). *Elements of Data Processing Mathematics, Third Edition.* Wilson T. Price, Charles Peselnick. Holt, Rinehart & Winston, 1987, xi + 670 pp, $26.50. [ISBN: 0-03-000178-1] A text for business students, it introduces them to such topics as modular arithmetic, bases 2, 8, and 16, flowcharts and sets, logical forms, basic algebra, simultaneous equations, and some linear programming techniques. It is clearly written and has lots of problems that relate the material to computers. MZ

Education, P. *Cognitive Processes in Mathematics.* Ed: John A. Sloboda, Don Rogers. Keele Cognition Seminars, V. 1. Oxford U Pr, 1987, x + 208

pp, $57.50. [ISBN: 0-19-852163-4] Ten papers from a 1985 conference on cognition. Contributors are psychologists, and mathematics is used as the medium in which to represent issues in cognition. Deals primarily with early competence in counting, computation, and simple algebra, and with representation of mathematical ideas. Note the price. MW

Education, P, L**.** *Twice As Less: Black English and the Performance of Black Students in Mathematics and Science.* Eleanor Wilson Orr. WW Norton, 1987, 240 pp, $15.95. [ISBN: 0-393-02392-3] A detailed analysis of how Black vernacular clashes with mathematical convention leading to incomprehensible confusion in which "half as large as" becomes "half as small as," and "one-third of" masquerades as "three times less." Performance of Black students in school mathematics is thus diminished by linguistic dissonance, with tragic consequences. Full of verbative examples; original; controversial; immensely important. LAS

Education, P, L. *Improving Indicators of the Quality of Science and Mathematics Education in Grades K-12.* Ed: Richard J. Murnane, Senta A. Raizen. National Academy Pr, 1988, ix + 220 pp, (P). [ISBN: 0-309-03740-9] An "indicator" is a statistic (e.g., test scores, enrollment data) used to monitor quality. This report from the National Research Council analyzes indicators in science and mathematics education of learning, of student behavior, of teaching quality, of curriculum, and of financial support, and makes recommendations concerning improvement in the reliability, quality, and usefulness of such indicators. LAS

Combinatorics, P, L. *Sphere Packings, Lattices and Groups.* J.H. Conway, N.J.A. Sloane. Grund. der math. Wissenschaften, V. 290. Springer-Verlag, 1988, xxvii + 663 pp, $87. [ISBN: 0-387-96617-X] What arrangement allows the largest number of baseballs to be packed in a box? This is the problem which motivates this book. Lattices and their automorphism groups enter the picture by considering the centers of the spheres in an arrangement. This book is a good comprehensive introduction to the sphere-packing problem and its relation to lattices, coding theory, and number theory. Many of the tables and constructions have not appeared before in print. LC

Combinatorics, P. *Lecture Notes in Mathematics-1278: Invariant Theory.* Ed: S.S. Koh. Springer-Verlag, 1987, 102 pp, $13.90 (P). [ISBN: 0-387-18360-4] A collection of seven papers authored by Pommerening, Formanek, Lascoux, Kung, Olver, Kempf, and Grosshans. SG

Combinatorics, T(16-17: 2), L. *Combinatorics of Experimental Design.* Anne Penfold Street, Deborah J. Street. Clarendon Pr, 1987, xiv + 400 pp, $27.50 (P); $55. [ISBN: 0-19-853255-5; 0-19-853256-3] Introduction to theory of designs with emphasis on aspects of interest to both combinatorialists and statisticians. The unifying concept is the pairwise balanced design. The main prerequisite is linear algebra; results from statistics, number theory, and finite field theory are introduced when needed. LC

Combinatorics, P. *Irregularities of Distribution.* József Beck, William W.L. Chen. Tracts in Math., V. 89. Cambridge U Pr, 1987, xiv + 294 pp, $54.50. [ISBN: 0-521-30792-9] A self-contained treatment of refinements and generalizations of the following theorem, first conjectured in 1935 by van der Corput: For any infinite sequence of real numbers in the unit interval, one can choose n and two subintervals of equal length such that the difference between the number of elements less than n in the subintervals is arbitrarily large. GG

Combinatorics, S(17), P. *Lecture Notes in Control and Information Sciences-94: Computational Complexity of Bilinear Forms.* Hari Krishna. Springer-Verlag, 1987, xv + 166 pp, $24.40 (P). [ISBN: 0-387-17661-6] A study of the relationship between the multiplicative complexity associated with the computation of a class of algebraic functions called bilinear forms and linear error-correcting codes. A new class of linear error-correcting codes is derived from the bilinear algorithms used for aperiodic convolution of certain sequences. CEC

Number Theory, L. *Geometry in Practice and Numbers in Theory.* Peter Hilton, Jean Pedersen. Mono. in Undergrad. Math., V. 16. Journal of Undergrad. Math. (Dept. of Math., Guilford Coll., Greensboro, NC 27410), 37 pp, $8.50 (P). Describes an algorithm for folding paper strips into certain polygons and investigates number theoretical questions which arise from this algorithm. Includes results from earlier papers along with some new results. An interesting exposition which could easily be used as the basis of an undergraduate independent study project. RH

Number Theory, S(15-16). *Fascinating Fractions.* N.M. Beskin. Transl: V.I. Kisin. Little Math. Lib. MIR (US Distr: Imported Pub), 1986, 87 pp, $2.95 (P). [ISBN: 0-8285-3195-1] A clearly-written exposition of the basic properties of continued fractions motivated by two interesting problems: approximations to π and the establishment of the Gregarian calendar. At this price a great bargain. SG

Linear Algebra, S(14). *Linear Algebra.* V.A. Ilyin, E.G. Poznyak. Transl: Irene Aleksanova. MIR (US Distr: Imported Pub), 1986, 285 pp, $10.95. [ISBN: 0-8285-3340-7] Very classical, theoretical treatment at higher than average level. Nine chapters cover matrices and determinants, linear spaces, systems of linear equations, Euclidean spaces, linear operators, iterative methods, bilinear and quadratic forms, tensors, elements of group theory. No exercises. GG

Group Theory, P. *Lecture Notes in Mathematics-1291: Correspondances de Howe sur un corps p-adique.* Colette Moeglin, Marie-France Vignéras, Jean-Loup Waldspurger. Springer-Verlag, 1987, vii + 163 pp, $17.30 (P). [ISBN: 0-387-18699-9]

Algebra, T(18). *Lectures on Artinian Rings.* Andor Kertész. Akademiai Kiado, 1987, 427 pp, $48.

[ISBN: 963-05-4309-5] A thorough treatment of the topic beginning with set theory to ring theory and moving through such topics as semi-simple rings, rings of quotients, group rings, and quasi-Frobenius rings. A modest number of exercises are included. SG

Algebra, T(12-13), P. *Algebra—aller Anfang ist leicht, 3. Auflage.* Herbert Kästner, Peter Göthner. BG Teubner, 1987, 155 pp, 8,40M (P). [ISBN: 3-222-00382-5] A gentle but careful introduction to abstract algebra for use in German high schools. JDB

Algebra, S(18), P. *Lecture Notes in Mathematics-1284: Konstruktive Galoistheorie.* B. Heinrich Matzat. Springer-Verlag, 1987, x + 286 pp, $25.80 (P). [ISBN: 0-387-18444-9] On the inverse problem of Galois theory for a field K, i.e., is every finite group the Galois group of some polynomial in $K[x]$? JD-B

Algebra, P. *Lecture Notes in Mathematics-1289: K-Theory, Arithmetic and Geometry.* Ed: Yu. I. Manin. Springer-Verlag, 1987, 399 pp, $39.40 (P). [ISBN: 0-387-18571-2] Nine articles originating in lectures given at the Manin Seminar at Moscow University over the years 1984-1986. GG

Calculus, C(13: 2), S*. *Computer Explorations in Calculus.* IBM PC or Apple II. K.D. Stroyan. Harcourt Brace Jovanovich, 1986, x + 278 pp, $13.95 (P). [ISBN: 0-15-512636-9] An introduction to numerical and graphical programming and a supplement to the first two semesters of calculus. Does not require previous experience with programming and programming technicalities are kept to a minimum. To be used concurrently with the appropriate topic in calculus. Programming is done in Basic. Lots of exercises. CEC

Calculus, T*(14: 1), L. *Mathematical Analysis: A Fundamental and Straightforward Approach.* David S.G. Stirling. Math. & Its Applic. Halsted Pr, 1987, 172 pp, $39.95. [ISBN: 0-470-20903-8] Contents often termed advanced calculus—lots of limits, some differentiation, series, and integration. Reasonable number of exercises, often with solution or hint in back. Frequently uses mathematical shorthand. Although it covers no material beyond first-year calculus, a good book for introducing rigor and proof. GG

Calculus, T(13: 1). *Calculus for the Managerial, Life, and Social Sciences.* S.T. Tan. Prindle, Weber & Schmidt, 1986, x + 525 pp. [ISBN: 0-87150-003-5] Suitable for one semester class. The text is very applications oriented; in fact, it has a list of applications from business and economics, social sciences, life sciences, and other areas. One quibble is that integration is not emphasized as area under curve. Nicely written. MZ

Real Analysis, S(18), P*. *Regular Variation.* N.H. Bingham, C.M. Goldie, J.L. Teugels. Encycl. of Math. & Its Applic., V. 27. Cambridge U Pr, 1987, xix + 491 pp, $75. [ISBN: 0-521-30787-2] "Comprehensive account of the theory and applications of regular variation. Concerned with the asymptotic behavior of a real function of a real variable x which is 'close' to a power of x." Well-motivated, clearly written, extensive bibliography, exercises at end of each chapter. BH

Complex Analysis, T(16: 1, 2), L. *Basic Complex Analysis, Second Edition.* Jerrold E. Marsden, Michael J. Hoffman. WH Freeman, 1987, xii + 604 pp. [ISBN: 0-7167-1814-6] "Substantial revision of the *First Edition* in detail, but not in spirit. Many passages have been rewritten for clarity." Added topics include integrals along continuous curves, normal families, Riemann mapping theorem, and functions of bounded variation, resulting in 130 page fattening of original (TR, April 1974). Main attraction remains numerous diagrams, examples, and exercises. BH

Differential Equations, T(15: 1), L. *Ordinary Differential Equations.* M.L. Krasnov. Transl: Irene Aleksanova. MIR (US Distr: Imported Pub), 1987, 150 pp, $4.95 (P). [ISBN: 0-8285-3444-6] This slim but packed volume is a translation of M.L. Krasnov's Russian book. There are five chapters: first-order differential equations, higher-order differential equations, systems of differential equations, stability theory, and some additional problems. There is a mix of theory (e.g., statements of unique existence of solutions to the Cauchy problem) and approximation concepts; major emphasis is on systems of differential equations. Includes a solid treatment of Lyapunov stability. Chapter exercises; suitable for a second course in differential equations. RSF

Differential Equations, P. *Lecture Notes in Mathematics-1285: Differential Equations and Mathematical Physics.* Ed: I.W. Knowles, Y. Saitō. Springer-Verlag, 1987, xvi + 499 pp, $48.50 (P). [ISBN: 0-387-18479-1] Papers from an international conference on differential equations and mathematical physics held at the University of Alabama, March 1986. MR

Partial Differential Equations, P. *Stochastiche Evolutionsgleichungen und deren Steuerung.* Wilfried Grecksch. Teubner-Texte zur Math., B. 98. BG Teubner, 1987, 212 pp, 22M (P). [ISBN: 3-322-00406-6] On stochastic evolution equations in the sense of Ito, and their control theory. JD-B

Numerical Analysis, P. *The State of the Art in Numerical Analysis.* Ed: A. Iserles, M.J.D. Powell. Inst. of Math. & Its Applic. Conf. Ser., V. 9. Clarendon Pr, 1987, xiv + 719 pp, $98. [ISBN: 0-19-853614-3] The proceedings of the joint IMA/SIAM conference held in April 1986 at the University of Birmingham. A collection of 23 papers covering the major developments of the last decade in numerical analysis. RH

Operator Theory, P. *Lecture Notes in Mathematics-1272: Commuting Nonselfadjoint Operators in Hilbert Space: Two Independent Studies.* Moshe S. Livšic, Leonid L. Waksman. Springer-Verlag, 1987, 115 pp, $13.10 (P). [ISBN: 0-387-18316-7] Two advanced discussions intended for specialists. AWR

Analysis, P, L. *N. Bourbaki, Elements of Mathematics: Topological Vector Spaces, Chapters 1-5.* Transl: H.G. Eggleston, S. Madan. Springer-Verlag,

1987, vii + 364 pp, $79. [ISBN: 0-387-13627-4] An English translation of the fifth book in the famous series. For those who have only heard about but never studied the Bourbaki style, this is as good a place as any to start. "In principle, it requires no particular knowledge of mathematics on the reader's part, but only a certain familiarity with mathematical reasoning and a certain capacity for abstract thought." Keep that in mind. AWR

Analysis, T(15-16: 1), L. *Methods of Summation.* Bertram Ross. Descartes Pr (2-13-10 Kaguike, Koriyama, Japan 963), 1987, ix + 127 pp, 4,000 yen (P). This little book is a collection of worked problems illustrating techniques in summing infinite series in closed form. Topics include the calculus of finite differences, the Psi function, fractional calculus and Fourier series. The prerequisites are some advanced calculus and knowledge of the Gamma and Psi functions. This should be of interest to applied mathematicians and engineers. LC

Analysis, P. *Seminar Analysis of the Karl-Weierstraß-Institute of Mathematics Academy of Sciences of the GDR 1985/86.* Ed: Bert-Wolfgang Schulze, Hans Triebel. Teubner-Texte zur Math., B. 96. BG Teubner, 1987, 322 pp, 33.50 DM (P). [ISBN: 3-322-00429-5] Collection of five papers in analysis touching topics such as vector-valued Sobolev spaces, multipliers in Besov-Triebel-Lizorkin spaces, non-linear partial differential equations, and scattering theory. LC

Differential Geometry, S(18), P. *Lecture Notes in Mathematics-1288: Generalized Analytic Functions on Riemann Surfaces.* Yuri L. Rodin. Springer-Verlag, 1987, v + 128 pp, $13.90 (P). [ISBN: 0-387-18572-0] Generalized analytic (pseudo-analytic) functions on Riemann surfaces arise as solutions to elliptic differential equations represented by the Carleman-Bers-Vekua equation $\bar{\partial}u + au + b\bar{u} = 0$. Properties such as the argument principle and the Liouville theorem hold for these functions. The author applies the theory to the Riemann-Roch theorem, to Abel's problem on functions with prescribed zeros and poles, and to some physical problems. GG

Geometry, T(15-17: 1), S, P, L. *Geometries and Groups.* V.V. Nikulin, I.R. Shafarevich. Transl: M. Reid. Universitext. Springer-Verlag, 1987, viii + 251 pp, $32.50 (P). [ISBN: 0-387-15281-4] A study of geometries that includes a classification based on the relation between their groups and the discrete groups of Euclidean space. The study reaches to the crystallographic groups, Lobachevskian geometry, as well as to interesting geometries on the torus. SS

Operations Research, S(18), P. *Theory of Linear and Integer Programming.* Alexander Schrijver. Ser. in Disc. Math. Wiley, 1986, xi + 471 pp, $71.95. [ISBN: 0-471-90854-1] A highly theoretical treatment of linear and integer programming suitable for graduate work. Discusses complexity theory, lattices and linear Diophantine equations. Extensive treatment of polyhedra, Khachiyan's method, and a brief discussion of Karmarkar's method. No exercises. SM

Operations Research, S(18), P. *Surveys in Game Theory and Related Topics.* Ed: H.J.M. Peters, O.J. Vrieze. CWI Tract, V. 39. Math Centrum, 1987, xii + 330 pp, Dfl. 50.40 (P). [ISBN: 90-6196-322-2] A set of thirteen survey papers concerning the current state-of-the-art in cooperative and non-cooperative game theory. SM

Operations Research, S(17), P. *Mathematical Modelling Methodology, Models and Micros.* J.S. Berry, *et al.* Math. & Its Applic. Halsted Pr, 1986, 318 pp, $89.95. [ISBN: 0-470-20717-5] A collection of 23 short articles on mathematical modeling. Directed at undergraduate and graduate school instructors. Includes a short chapter on using modeling techniques in calculus instruction. Emphasis on models and software suitable for implementation on microcomputers. SM

Operations Research, T(17-18: 1), P. *Team Theory.* K.H. Kim, F.W. Roush. Math. & Its Applic. Halsted Pr, 1987, ix + 246 pp, $54.95. [ISBN: 0-470-20848-1] A "team" is a group of "persons" with identical interests but independent information and actions. The field has similarities to game theory. Cooperation is the underlying behavioral assumption. Many exercises. Extensive bibliography. SM

Dynamical Systems, P. *Periodic Solutions of Hamiltonian Systems and Related Topics.* Ed: P.H. Rabinowitz, *et al.* NATO ASI Ser. C, V. 209. D Reidel (US Distr: Kluwer Academic), 1987, xi + 284 pp, $59. [ISBN: 90-277-2553-5] Proceedings of a NATO Advanced Research Workshop held in Il Ciocco, Italy, in October 1986. AWR

Control Theory, P. *Flow Control of Congested Networks.* Ed: Amedeo R. Odoni, Lucio Bianco, Giorgio Szegő. NATO ASI Ser. F, V. 38. Springer-Verlag, 1987, xii + 355 pp, $69.50. [ISBN: 0-387-18398-1] A collection of papers presented during the NATO Advanced Research Workshop on flow control of congested networks, the case of data processing and transportation, held in Capri, Italy, October 12-18, 1986. Papers focus on current research in flow control methodologies and their application to congestion reduction within data communication networks, transportation networks, and air traffic control systems. PS

Systems Theory, P. *Information Systems: Failure Analysis.* Ed: John A. Wise, Anthony Debons. NATO ASI Ser. F, V. 32. Springer-Verlag, 1987, xv + 338 pp, $59.50. [ISBN: 0-387-17800-7] Twenty-nine papers concerning the failure of information systems. The material is intended to provide guidelines for both the design of information systems and the prevention of future failures. Discusses methodologies for the investigation of information system failures. Examples of well-known information systems failures are the accident at Three Mile Island and the space shuttle Challenger explosion. SM

Systems Theory, T(17: 1), P. *Introduction to Queueing Networks.* E. Gelenbe, G. Pujolle. Transl: J.C.C. Nelson. Wiley, 1987, xiv + 177 pp,

$29.95. [ISBN: 0-471-90464-3] A textbook introducing queueing networks. Begins with standard simple queues, gradually increasing generality. Examples from data transmission networks. A few problems, most with solutions. SM

Probability, P. *Lecture Notes in Mathematics-1247: Séminaire de Probabilités XXI.* Ed: J. Azéma, P.A. Meyer, M. Yor. Springer-Verlag, 1987, iv + 579 pp, $54.40 (P). [ISBN: 0-387-17768-X] Papers collected from Institut de Mathematiques, Universite de Strasbourg. Over half of the papers are in French. Topics covered include homogeneous chaos, construction of the Malliavin operator on Poisson space, stationary Markov spaces, approximation of predictable characteristics of processes with filtration. MZ

Stochastic Processes, P. *Stochastic Differential Systems, Stochastic Control Theory and Applications.* Ed: Wendell Fleming, Pierre-Louis Lions. IMA, V. 10. Springer-Verlag, 1988, xiii + 609 pp, $49.80. [ISBN: 0-387-96641-2] Proceedings of a workshop held at the Institute for Mathematics and Its Applications (University of Minnesota), June 1986. Topics addressed include stochastic scheduling, applications to problems of computer networks, and stochastic gradient algorithms. MR

Stochastic Processes, T(17: 1), P. *Correlation Theory of Stationary and Related Random Functions.* A.M. Yaglom. Ser. in Stat. Springer-Verlag. *Volume I: Basic Results,* 1986, xiv + 526 pp, $58 [ISBN: 0-387-96268-9]; *Volume II: Supplementary Notes and References,* 1987, vii + 258 pp, $56.50. [ISBN: 0-387-96331-6] *Volume I* presents the "most important part of the theory of stationary and related random functions dealing only with first and second moments of these functions," emphasizing applied aspects of the theory rather than proofs. No exercises. *Volume II* (note price) contains a diverse set of notes, ranging from historical comments to detailed proofs, together with a more complete bibliography of both theoretical and applied references. No exercises. RSK

Stochastic Processes, T(17-18: 1). *Introduction to Random Processes.* Yuriĭ A. Rozanov. Transl: Birgit Röthinger. Springer-Verlag, 1987, viii + 117 pp, $35. [ISBN: 0-387-17874-0] Begins with a study of homogeneous Markov processes with a countable number of states, followed by an outline of the foundations of stochastic analysis. Studies the problem: "Given some relatively simple characteristics of a process, compute the probability of another event which may be very complicated; or estimate a random variable which is related to the behavior of the process." Introduces concepts by referring to mathematical models that are simple, but important in applications. RH

Statistics, P. *Modern Statistical Selection, Part 1.* Ed: M. Haseeb Rizvi, *et al.* Amer. J. of Math. & Manag. Sci., V. 5, Nos. 3 & 4. American Sciences Pr, 1985, 201 pp, $49.75 (P). A collection of papers discussing the development of statistical ranking and selection. SM

Statistics, T(17-18), P. *Tensor Methods in Statistics.* Peter McCullagh. Chapman & Hall, 1987, xv + 285 pp, $35. [ISBN: 0-412-27480-9] Text explains how index notation and tensors can be utilized to simplify calculations in statistics. Topics include cumulants, Edgeworth series, saddle-point approximation, and likelihood functions. LC

Computer Literacy, S*, L.** *Portraits in Silicon.* Robert Slater. MIT Pr, 1987, xiv + 374 pp, $24.95. [ISBN: 0-262-19262-4] 31 biographical vignettes tracing the opportunities, frustrations, motivations, and accomplishments of some of the pioneers of the computer age (Turing, Perot, Noyce, Cray, Hopper, Kemeny, Jobs, Knuth, ...). Engagingly written in clear yet authentic terms; based largely on personal interviews by the author, a professional reporter and biographer. LAS

Programming, T*(13: 1), S, L. *Pascal: An Introduction to the Art and Science of Programming, Second Edition.* Walter J. Savitch. Struct. Prog. Benjamin/Cummings, 1987, xxv + 769 pp, $27.95 (P). [ISBN: 0-8053-8388-3] This book has been thoroughly rewritten (*First Edition,* TR, November 1986). The basic topic groupings and the early and continued use of procedures remain even though entire chapters have disappeared and have been replaced with new ones. Specifically, expanded topics include problem solving, debugging, programming techniques, arrays, and recursion. CEC

Languages, T(17-18: 1), S, P, L. *Understanding Language: Man or Machine.* John A. Moyne. Found. of Comput. Sci. Plenum Pr, 1985, xvi + 357 pp, $49.50. [ISBN: 0-306-41970-X] A graduate text for computer scientists and linguists interested in natural language processing in artificial intelligence. Focus on linguistic, psycholinguistic prerequisites, issues in human linguistic behavior, rather than on extant models and algorithms. Preliminaries; grammars and parsing; semantics; comments on artificial intelligence approaches; lexical processing; syntactic processing. Self-contained, no specific prerequisites; exercises, attention to references. RB

Algorithms, S. *Algorithms for Games.* G.M. Adelson-Velsky, V.L. Arlazarov, M.V. Donskoy. Springer-Verlag, 1988, x + 197 pp, $59.50. [ISBN: 0-387-96629-3] Two-person, zero-sum games with complete information; chess is the model game; α-β pruning, heuristics, analogical reasoning. Preliminary results on a probabilistic approach. Translation of a 1978 Russian publication. RWN

Computer Systems, P. *Distributed Operating Systems: Theory and Practice.* Ed: Yakup Paker, Jean-Pierre Banatre, Müslim Bozyiğit. NATO ASI Ser. F, V. 28. Springer-Verlag, 1987, x + 379 pp, $77.50. [ISBN: 0-387-17699-3] Proceedings of NATO Advanced Study Institute on distributed operating systems, Altinyunus, Turkey, August 1986. Fourteen papers concerning formal aspects of concurrent systems; design issues for distributed operating systems; hardware support for distributed operating systems;

case studies of existing research and commercial implementations. RB

Computer Systems, S(14-15). *Introduction to Real-time Software Design.* S.T. Allworth, R.N. Zobel. Springer-Verlag, 1987, xiii + 287 pp, $28 (P). [ISBN: 0-387-91307-6] This *Second Edition* updates and expands on the *First Edition,* adding interfaces, real-time languages, distributed systems, and signal processing. It is divided into three parts: the real-time virtual machine, designing the processes, and designing and measuring the system. Each chapter has a summary and a list of concepts. It presents a structured design of software, concentrating on basic principles, and is language independent. Includes a list of references, a bibliography, and a comprehensive glossary. RSF

Computer Science, P. *Lecture Notes in Computer Science-289: ESEC '87.* Ed: H.K. Nichols, D. Simpson. Springer-Verlag, 1987, xii + 404 pp, $33.30 (P). [ISBN: 0-387-18712-X] Proceedings of the first European software engineering conference held at Strasbourg in September 1987. Forty papers covering a wide range of topics from formal methods to application systems. PS

Applications, T??, L??. *Modern Methods of Music Analysis Using Computers.* R.M. Mason. Schoolhouse Pr (46 Mountain View Dr., Peterborough, NH 03458), 1985, 299 pp, $44.35. [ISBN: 0-9615669-0-6] "Using essentially only the elementary mathematics of vectors and complex numbers," this publication is offered as a text for a course in music analysis by computers or a graduate seminar in "analytic tonality," simultaneously as a profesional-level reference for computer scientists interested in musical applications, and as a tutorial-level exposition for the advanced music student and the serious computer hobbyist. Algorithms expressed in Reverse Polish Notation. RB

Applications, P. *Modelling and Analysis in Arms Control.* Ed: Rudolf Avenhaus, Reiner K. Huber, John D. Kettelle. NATO ASI Ser. F, V. 26. Springer-Verlag, 1986, viii + 488 pp, $96.50. [ISBN: 0-387-17174-6] Analyzes arms control from a game theoretic/dynamical systems point of view. Four parts: history, pre-arms control assessment of options, negotiations, verification. SM

Applications (Biological Science), P. *Lecture Notes in Biomathematics-73: Applications of Control Theory in Ecology.* Ed: Y. Cohen. Springer-Verlag, 1987, vii + 101 pp, $16.30 (P). [ISBN: 0-387-18104-0] Proceedings of the symposium on optimal control theory held at the State University of New York, Syracuse, on August 10-16, 1986. Five papers using the techniques of control theory to study biological systems. SM

Applications (Cognitive Science), P. *Computational Models of Learning.* Ed: Leonard Bolc. Symbolic Computat. Springer-Verlag, 1987, ix + 208 pp, $45. [ISBN: 0-387-16318-2] A collection of essays on selected problems in machine learning. Includes learning strategies for expert systems, empirical discovery of numeric laws, analysis of heuristic learning, breakdown of goals into subgoals, induction and concept formation, theory of linguistic and cognitive development. References. Subject Index. RJA

Applications (Engineering), P. *Non-Classical Continuum Mechanics.* Ed: R.J. Knops, A.A. Lacey. London Math. Soc. Lect. Note Ser., V. 122. Cambridge U Pr, 1987, xii + 335 pp, $37.50 (P). [ISBN: 0-521-34935-4] Twenty papers on practical and abstract aspects of continuum mechanics, including applied mathematical techniques, theoretical mechanics, and structural and aeronautical engineering. MR

Applications (Engineering), T(17). *Computational and Applied Mathematics for Engineering Analysis.* A.S. Cakmak, J.F. Botha, W.G. Gray. Springer-Verlag, 1987, viii + 380 pp, $69. [ISBN: 0-387-17505-9] Presents analytical, approximate, and numerical methods for the solution of engineering problems, assuming elementary background in linear algebra, vector analysis, differential equations, and complex analysis. Problems at end of each chapter. Many typographical errors. BH

Applications (Information Theory), S*(14-16), L*. *A Diary on Information Theory.* Alfréd Rényi. Appl. Prob. & Stat. Wiley, 1987, ix + 125 pp, $34.95. [ISBN: 0-471-90971-8] Reprint by Wiley of the 1984 English Akademiai Kiado edition (TR, October 1985). An imaginary rendering of a student's first encounter with lectures on information theory. LAS

Applications (Medicine), P. *Mathematical Modelling in Biomedicine: Optimal Control of Biomedical Systems.* Y. Cherruault. Math. & Its Applic. D Reidel (US Distr: Kluwer Academic), 1985, xviii + 258 pp, $49.50. [ISBN: 90-277-2149-1] Serves as an introduction to the mathematical modeling of biological systems. Includes sections on control theory, integral equations and partial differential equations. Also presents brief discussion of open problems in biomathematics. Bibliography. SM

Reviewers

RJA: Richard J. Allen, St. Olaf; DFA: David F. Appleyard, Carleton; RB: Richard Brown, Carleton; LB-E: Lisa Brown-Ewing, Macalester; JNC: Judith N. Cederberg, St. Olaf; LC: Laura Chihara, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; RSF: Robert S. Fisk, St. Olaf; SG: Steven Galovich, Carleton; GG: George Gilbert, St. Olaf; BH: Bruce Hanson, St. Olaf; RH: Rhonda Hatcher, St. Olaf; RK: Roger Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; AM: Alan Magnuson, St. Olaf; SM: Steve McKelvey, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; MR: Matthew Richey, St. Olaf; AWR: A. Wayne Roberts, Macalester; GMS: G. Michael Schneider, Macalester; PS: Paul Schoening, St. Olaf; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MS: Myriam Steinback, Macalester; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; LW: Lynne Walling, St. Olaf; FLW: Frank L. Wolf, Carleton; MZ: Michael Zielinski, St. Olaf.

Random Walks and Electric Networks,

by J. Laurie Snell and Peter Doyle

xiii + 159 pages. Hardbound
List: $25.00 MAA Member: $19.00

In this newest addition to the Carus Mathematical Monographs, the authors examine the relationship between elementary electric network theory and random walks, at a level which can be appreciated by the able college student. We are indebted to them for presenting this interplay between probability theory and physics in so readable and concise a fashion.

Central to the book is Polya's beautiful theorem that a random walker on an infinite street network in d-dimensional space is bound to return to the starting point when d = 2, but has a positive probability of escaping to infinity without returning to the starting point when d = 3. The authors interpret this theorem as a statement about electric networks, and then prove the theorem using techniques from classical electrical theory. The techniques referred to go back to Lord Rayleigh who introduced them in connection with an investigation of musical instruments.

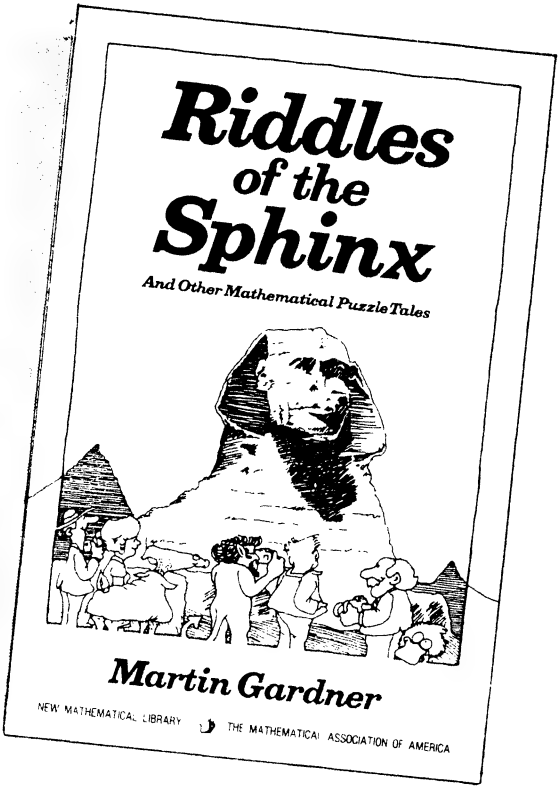
In Part I the authors restrict themselves to the study of random walks on finite networks, establishing the connection between the electrical concepts of current and voltage and corresponding descriptive quantities of random walks regarded as finite state Markov chains. Part II deals with the idea of random walks on infinite networks.

Table of Contents

Order From:
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

Riddles of the Sphinx

and other mathematical puzzle tales.



by Martin Gardner
Volume 32 in the New Mathematical Library
184 pp., Paper. ISBN-0-88285-632-8
List: $14.50   MAA Member: $12.50

Martin Gardner has charmed readers for over fifty years with his delightful books and articles on science. He is best known for the popular column, *"Mathematical Games,"* which appeared in *Scientific American* for twenty-five years. Generations of scientists and mathematicians have been inspired by his writing and the MAA is proud to include his name in its list of authors.

This book was drawn from Gardner's column in *Isaac Asimov's Science Fiction Magazine.* The riddles presented here incorporate the responses of his initial readers, along with additions suggested by the editors of the New Mathematical Library. Each chapter (riddle) poses a problem answered in the First Answers section. The solution in turn raises another problem that is solved in the Second Answers section. This may suggest a third question and in several instances there is a fourth. Gardner draws us from questions to answers always presenting us with new riddles—some as yet unanswered. There are 125 different pieces altogether.

Solving these riddles is not simply a matter of logic and calculation, although these play a role. Luck and inspiration are factors as well, so beginners and experts alike may profitably exercise their wits on Gardner's problems, whose subjects range from geometry to word play to questions relating to physics and geology.

We guarantee that you will solve some of the riddles, be stumped by others, and be amused by almost all of the stories and settings that Gardner has devised to raise these questions.

Order from:
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 95, Number 6 June-July 1988

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Dept. of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Shorter articles about the teaching of mathematics should be sent to STAN WAGON and JOAN HUTCHINSON, Department of Mathematics, Smith College, Northampton, MA 01063. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to ANITA E. SOLOW, Dept. of Mathematics, Grinnell College, Grinnell, IA 50112; in geometry or topology to DENNIS DETURCK, Dept. of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to DAVID J. HALLENBECK, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Proposed problems (three copies) and solutions (two copies), both elementary and advanced to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, University of Calgary, Alberta, Canada T2N 1N4.

Reprint Permission: A. B. WILLCOX, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Back Issues: P. and H. BLISS Co., Middletown, CT 06457.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $26 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $29. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $90 per year.

Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Assocation of America, 1529 Eighteenth Street, N.W., Washington, DC 20077-9564.

PRINTED IN THE UNITED STATES OF AMERICA

Growth in Gaussian Elimination

JANE DAY AND BRIAN PETERSON, *San Jose State University*

JANE DAY received her degrees at the University of Florida, the B.A. in 1958 and Ph.D. in 1964. After two postdoctoral years at Florida, she visited the Institute for Advanced Study in 1966-67, as an AAUW Fellow; from 1967-81, she taught at the College of Notre Dame in Belmont, CA, and has been at San Jose State University since.

BRIAN PETERSON received his B.A. at San Jose State University in 1972 and his Ph.D. at UC Berkeley in 1976. He taught for two years at Rutgers University and has been at San Jose State University since 1978.

Dedicated to the memory of James H. Wilkinson, 1919–1986

1. Introduction. This article is about a conjecture made by J. H. Wilkinson during some of his early work in numerical analysis (see Wilkinson [1961]). We will discuss what has been proved about the problem, extend some of those results, and provide examples that destroy some seemingly natural conjectures about his assertion and raise some interesting new questions.

Wilkinson's conjecture is very intriguing—easy to state, soon believed, and apparently very difficult to resolve. Fix a positive integer n, let $A = [a_{ij}]$ be an $n \times n$ matrix and reduce A to upper triangular form by Gaussian elimination with complete pivoting (precisely defined in (2.1)). Let $A^{(k)} = [a_{ij}^{(k)}]$ denote the matrix obtained after the first k pivoting operations, so $A^{(n-1)}$ is the final upper triangular matrix. A diagonal entry of that final matrix will be called a *pivot*. Wilkinson conjectured that $g(A) \leqslant n$ for all real $n \times n$ matrices A, where

$$g(A) = \left(\max_{i,j,k}\left|a_{ij}^{(k)}\right|\right)\Big/\left(\max_{i,j}|a_{ij}|\right).$$

It follows easily from the definition of complete pivoting that $g(A)$ always occurs at some pivot.

We are interested here only in real matrices; in fact, for an $n \times n$ complex matrix A, $g(A)$ can exceed n (Tornheim [1965], [1970]). Assume henceforth that all matrices are real.

It is known that the conjectured bound for growth will be sharp if indeed the conjecture is true, because an $n \times n$ Hadamard matrix A has its last pivot equal to $\pm n$ (see Section 5). Furthermore, there is considerable evidence that, when an

$n \times n$ Hadamard matrix is reduced by Gaussian elimination with complete pivoting, no pivot before the last one can have magnitude exceeding n: earlier authors showed that the magnitudes of the last three pivots are $n/2$, $n/2$ and n; and we extend that in Section 5, proving that none of the last six pivots can have magnitude greater than n. However, unlikely as it seems, so far as is known now some still earlier pivot might have magnitude greater than n.

In Section 5 we will also establish values for the first five pivots of any Hadamard matrix reduced by complete pivoting, and will show that $g(A)$ does equal n for a certain class of $n \times n$ Hadamard matrices A.

It will be useful to have the following notation:

$$g(n) = \sup\{ g(A) | A \text{ an } n \times n \text{ matrix}\}.$$

In this notation, Wilkinson's conjecture says that $g(n) \leqslant n$ for all n. We will call the problem of determining $g(n)$ the *growth problem*.

Section 2 contains essential classical background material, various algebraic and geometric interpretations of Gaussian Elimination with complete pivoting, and results of ours which justify use of the nonlinear optimization software package NPSOL [Gill et al., 1983]. Our theorems include showing that the bound $g(n)$ is achieved for some matrix in a certain set of $n \times n$ matrices and that it occurs at the last pivot of that matrix.

Section 6 is a report on the examples we obtained using NPSOL, and the new questions they suggest. We attempted to achieve the bound $g(n)$ for $n \leqslant 7$. We also used NPSOL to test the conjecture that $g(n)$ will occur in an orthogonal matrix for all n. This seems plausible at first because $g(n)$ does appear to occur in Hadamard matrices in the sizes where they exist, and the natural analogue of Hadamard matrices in other sizes would seem to be orthogonal matrices. However, the results indicate that conjecture is almost certainly false for $n = 5$ and $n = 7$.

In Section 3 we will discuss the best general bound for $g(n)$ known so far,

$$g(n) \leqslant \left[n\, 2\, 3^{1/2} \cdots n^{1/n-1}\right]^{1/2}.$$

This bound is due to Wilkinson [1961] and is much larger than his conjectured n; we will explain why his method of proof could not yield a considerably better bound.

To evaluate $g(n)$ directly is a constrained optimization problem with a nonlinear objective function and many nonlinear constraints. (See Section 6.) While it is easy to see that $g(1) = 1$ and $g(2) = 2$, for all $n > 2$, even $n = 3$ or 4, it seems quite difficult to establish a value or close bound. Cryer [1968], Cohen [1974], and Tornheim [1964], [1965], [1969], [1970] proved that $g(3) = 2.25$ and $g(4) = 4$, and Tornheim also showed that $g(5) \leqslant 4\ 17/18$. Their methods are algebraic, with many cases, and would be prohibitively tedious for large n. Tornheim invokes the Kuhn-Tucker theory for nonlinear constrained optimization in one proof, but the argument is still very involved.

In Section 4, we give our own proof that $g(3) = 2.25$, which uses geometry to make various cases and inequalities easier to see. Unfortunately this geometric viewpoint does not appear adaptable to larger n either. Apparently the value of $g(n)$ has not been shown to be bounded by n for any single value of $n \geqslant 6$.

One of the curious frustrations of the growth problem is that it is quite difficult to construct any examples of $n \times n$ matrices A other than Hadamards for which $g(A)$ is even close to n. Wilkinson has remarked that in real-world problems, $g(A)$

has never been observed to be very large (p. 214, Wilkinson [1965]). C. W. Cryer [1968] did numerical experiments in which he computed $g(A)$, doing complete pivoting on $n \times n$ matrices A with entries chosen randomly from the interval $[-1,1]$ and for sizes up to $n = 8$. He had to generate over 50,000 3×3 examples before finding one with $g(A) > 2$; and although it is true that $g(5) > 4$, $g(8) \geqslant 8$, and $g(n)$ is nondecreasing, every example he obtained through $n = 8$ had $g(A) \leqslant 2.8348$.

To complete the introduction, we will indicate briefly why complete pivoting matters in numerical analysis, and how $g(A)$ gets involved. To solve a linear system $A\underline{x} = \underline{b}$ using the usual Gaussian elimination algorithm (not doing complete pivoting), one simply chooses the first available nonzero entry in rows and columns $k, \ldots, n$ for the kth pivot, exchanging rows and columns if necessary to put that entry into the (k, k) position. When Wilkinson invented backward error analysis for this algorithm, $g(A)$ appeared in the bound for relative error. To sketch how, suppose that t-digit rounding arithmetic in base β is used, and assume that A is invertible. Let $\underline{x}$ denote the theoretically correct solution and $\underline{y}$ the calculated solution. Wilkinson proved that there is an error matrix E for which $(A + E)\underline{y} = \underline{b} + \underline{e}$, and each $|e_{ij}|$ and $|e_i|$ is bounded above by $cg(A)$ where c is about β^{1-t}. Thus $\underline{y}$ is the exact solution to a nearby problem if $g(A)$ is not large. Using any matrix norm consistent with matrix product, such as the ∞-norm, it is possible to show that the relative error in $\underline{y}$ is bounded thus:

$$\frac{\|\underline{y} - \underline{x}\|}{\|\underline{x}\|} \leqslant \|A^{-1}\|\,\|A\|\,\|E\|d,$$

where d is not alarmingly large and $\|E\|$ depends of course on the size of the entries in E, hence on $g(A)$.

Now complete pivoting means this: after $k - 1$ pivot steps, choose the kth pivot to have maximum magnitude among all entires in the last $n - k + 1$ rows and columns. Such a choice minimizes round-off error at the kth pivot step, which seems sensible. Thus the question of how large $g(A)$ can be when complete pivoting is done became of interest. Details can be found in (Sec. 24–27, Wilkinson [1965]) or (Sec. 4.2–4.4, Golub and Van Loan [1985]).

Many decades of computation have passed since that error analysis was first done; all experience has shown that $\|E\|$ is not large, and that when relative error is unacceptable for a problem $A\underline{x} = \underline{b}$, it is not because of growth but because the condition number $\|A^{-1}\|\,\|A\|$ of A is large. In fact, in practice this is true even if only partial pivoting is done (see Definition (2.1)). Thus the question of finding a better bound for $g(n)$ has not been an issue in numerical analysis for many years.* Yet the mathematical problem remains tantalizing. It led these authors to explore a number of interesting ideas, including some lovely old determinant theory which is not widely known today. What follows is a report of what we learned, and some new results about $g(n)$.

2. Notation and basic results. The following establishes precise notation for the algorithm commonly called complete pivoting, and for the growth associated with it.

*However, see Trefethen [1985] for some provocative questions about this and other matters long considered settled.

(2.1) *Definitions*. Let $B[k]$ denote the lower right $k \times k$ principal submatrix of a matrix B.

Let $A = [a_{ij}]$ denote a nonzero $n \times n$ real matrix. To reduce A by *Gaussian Elimination with complete pivoting* (GECP) is to perform steps (a)–(c):

(a) Search column 1 from row 1 to n, then columns $2, \ldots, n$ in the same way, for an entry with maximum magnitude. Using the first such a_{ij} found, exchange rows 1 and i and columns 1 and j if necessary to put it into the $(1,1)$ position. Let $A^{(0)} = [a_{ij}^{(0)}]$ denote the new matrix.

(b) Pivot on $a_{11}^{(0)}$—that is, add multiples of row 1 to the rows below it to create zeroes in their column 1 positions. Let $A^{(1)} = [a_{ij}^{(1)}]$ be the new matrix obtained.

(c) In the same way for each $k = 2, \ldots, n-1$, search $A^{(k-1)}[n-k+1]$ to find the maximum magnitude of an entry, and exchange rows and columns of A if necessary to put the first such entry found into the (k, k) position. If it is not zero, pivot on it, to obtain $A^{(k)}$; otherwise, let $A^{(k)} = A^{(k-1)}$. Thus the final upper triangular matrix can be denoted thus:

$$A^{(n-1)} = \begin{bmatrix} a_{11}^{(0)} & a_{12}^{(0)} & \cdots & a_{1n}^{(0)} \\ 0 & a_{22}^{(1)} & \cdots & a_{2n}^{(1)} \\ 0 & 0 & \cdots & \cdot \\ 0 & 0 & \cdots & a_{nn}^{(n-1)} \end{bmatrix}.$$

As mentioned in Section 1, the diagonal entries $a_{kk}^{(k-1)}$ are called the *pivots*. It is true that A is invertible if and only if every pivot is nonzero.*

Using this notation,

$$g(A) = \max_{i,j,k} \left|a_{ij}^{(k)}\right| \Big/ \left|a_{11}^{(0)}\right|.$$

We will call $g(A)$ the *growth* associated with GECP on A.

By the definition of GECP, the magnitude of some pivot will be $g(A)$—but it need not be the last pivot. However, we will prove below that there is an $n \times n$ matrix on which $g(A)$ has the maximum value $g(n)$ and for which $g(A)$ does occur at the last pivot, $a_{nn}^{(n-1)}$.

If one knew ahead of time what row and column exchanges would occur during GECP and did them to A first, in the same order, then it is true that GECP on this altered matrix would not actually require any exchanges, and the final matrix would be the same $A^{(n-1)}$. So to discuss growth we can restrict attention to matrices with the property that no exchanges are actually needed during GECP. We call such matrices *completely pivoted* (CP).

The following notation and proposition will be used throughout.

(2.2) *Definition*. Let $A(i_1 \cdots i_p | j_1 \cdots j_p)$ denote the determinant of the $p \times p$ submatrix of A obtained from the intersection of rows $i_1, \ldots, i_p$ with columns $j_1, \ldots, j_p$. When the two sets of indices are the same, $A(i_1 \cdots i_p)$ will abbreviate.

*If in step (c) one searches only the kth column from row k to row n for the first entry of maximum magnitude to use as the kth pivot, that is called *Gaussian Elimination with partial pivoting*. If one uses the first nonzero entry found instead of the first one of maximum magnitude, that is, *usual Gaussian Elimination*.

Let adj A denote the matrix with $(-1)^{i+j}A(1\cdots\hat{\imath}\cdots n|1\cdots\hat{\jmath}\cdots n)$ as its (j, i) entry, where $\hat{\imath}$ means "omit i." As is well known, $A^{-1}=(1/\det A)(\operatorname{adj} A)$ if A is invertible.

(2.3) PROPOSITION (p. 26, Gantmacher [1959]). *Suppose A is CP, invertible, GECP is done on A, and $1 \leqslant k < n$. Then if $k < i, j$,*

$$a_{ij}^{(k)} = A(1\cdots ki|1\cdots kj)/A(1\cdots k).$$

Thus the kth pivot $a_{kk}^{(k-1)}$ is $1/x$ where x is the (k, k) entry of the inverse of the $k\times k$ leading principal submatrix of A; in particular, the last pivot $a_{nn}^{(n-1)}$ is the reciprocal of an entry of A^{-1}.

Proof. The determinant of a matrix is unchanged by adding multiples of a row to other rows. Because of this, and because no exchanges will be done during GECP, the $k\times k$ leading principal submatrices of A and $A^{(n-1)}$ have the same determinant; that is, $A(1\cdots k)=a_{11}^{(0)}a_{22}^{(1)}\cdots a_{kk}^{(k-1)}$. Since A is invertible and CP, $A(1\ldots k)\neq 0$ and this is the denominator of the quotient in the theorem.

Now adjoin $k+1$ entries of row i and column j to get the submatrix of A whose determinant is $A(1\cdots ki|1\cdots kj)$. Just as above, this equals $a_{11}^{(0)}a_{22}^{(1)}\cdots a_{kk}^{(k-1)}a_{ij}^{(k)}$. This is the numerator of the quotient in the theorem, hence the equality is proved.

Finally, if B denotes the $k\times k$ leading principal submatrix of A, then the (k,k) entry of B^{-1} is $B(1\cdots(k-1))/B(1\cdots k)=A(1\cdots(k-1))/A(1\cdots k)$ and this equals $1/a_{kk}^{(k-1)}$ by the preceding. □

The above result shows that GECP can be interpreted thus: having chosen the first $k-1$ rows and columns, so that the $(k-1)\times(k-1)$ leading principal subdeterminant is determined, choose the kth row and column so that the $k\times k$ leading principal subdeterminant will have maximum possible magnitude.

An alternative interpretation is this: having chosen the first $k-1$ rows and columns, choose the kth pivot so that $|\det A^{(k)}[n-k]|$ is minimal—for $|a_{kk}^{(k-1)}|$ is maximal among the entries of $A^{(k-1)}[n-k+1]$ and

$$A^{(k-1)}[n-k+1]=\left[\begin{array}{c|c} a_{kk}^{(k-1)} & \cdots\ a_{kn}^{(k-1)} \\ \cdots\cdot & \cdots\cdots\cdots \\ 0 & A^{(k)}[n-k] \end{array}\right].$$

For geometric interpretations of GECP, recall that the magnitude of the determinant of a $k\times k$ matrix is the k-volume of the parallelepiped in R^k which is spanned by its rows. The obvious geometric interpretation of GECP then is that it means "after the $(k-1)$st pivot step, choose the kth row and column so as to maximize the k-volume of the parallelepiped spanned by the rows of the $k\times k$ leading principal submatrix."

Alternatively, GECP also means "minimize the $(n-k)$-volume of the projected $(n-k)$-dimensional solid at the kth pivot step": for pivoting on $a_{kk}^{(k-1)}$ amounts to projecting the parallelepiped spanned by the rows of $A^{(k-1)}[n-k+1]$, by vectors parallel to its first row, into the parallelepiped spanned by the rows of $A^{(k)}[n-k]$, which is of one lower dimension.

In Proposition (2.5) below, we give what appears to be a new algebraic interpretation of GECP. That theorem, and much else that follows in this paper, depends on a formula from Gantmacher [1959], which we state in the next Proposition. While the

formula is important in what follows here, its proof is somewhat lengthy and would not add insight into our applications. Briefly, the formula follows from two classic theorems, one due to A. Cauchy and J. Binet about compound matrices, and the general Laplace Expansion for $\det A$. Both of these classic results date from the early 19th century, and they are nicely presented in Sec. I.4, Gantmacher [1959]; also see Marshall and Olkin [1979].

(2.4) PROPOSITION (p. 21, Gantmacher [1959]). *Let A be an $n \times n$ invertible matrix and let $k < n$. If $i_1 < i_2 < \cdots < i_k$ is an ordered subset of $\{1, 2, \ldots, n\}$, let $i_1' < i_2' < \cdots < i_{n-k}'$ denote its ordered complement. Then for any k rows and columns,*

$$A^{-1}(i_1 \cdots i_k | j_1 \cdots j_k) = \frac{(-1)^m}{\det A} A(j_1' \cdots j_{n-k}' | i_1' \cdots i_{n-k}'),$$

where $m = \sum_{r=1}^{k} (i_r + j_r)$. □

Note (2.4) is a strong generalization of the much better known fact that $A^{-1} = (1/\det A)(\operatorname{adj} A)$.

(2.5) PROPOSITION. *Let A be an $n \times n$ invertible matrix and $1 \leqslant k < n$. Do k steps of usual Gaussian Elimination with no row or column exchanges. Then the unreduced part of $A^{(k)}$ can be described thus:*

$$A^{(k)}[n-k] = \left(A^{-1}[n-k]\right)^{-1}.$$

So having the rows and columns of A arranged to make it CP means this: for each $k = 1, \ldots, n-1$ the magnitude of the determinant of $A^{-1}[n-k]$ cannot be increased by rearranging the last $n-k+1$ rows and columns of A.

Proof. Let $i, j > k$ so that $a_{ij}^{(k)}$ is an arbitrary entry of $A^{(k)}[n-k]$. Then by (2.3) and (2.4) we have

$$\begin{aligned} a_{ij}^{(k)} &= A(1 \cdots k\,i | 1 \cdots k\,j)/A(1 \cdots k) \\ &= \frac{(-1)^{i+j}(\det A)A^{-1}((k+1) \cdots \hat{j} \cdots n | (k+1) \cdots \hat{\imath} \cdots n)}{A(1 \cdots k)}. \end{aligned}$$

Also by (2.4),

$$\begin{aligned} \det\left(A^{-1}[n-k]\right) &= A^{-1}((k+1) \cdots n) \\ &= A(1 \cdots k)/(\det A). \end{aligned}$$

Thus we have

$$a_{ij}^{(k)} = \frac{(-1)^{i+j}}{\det\left(A^{-1}[n-k]\right)} A^{-1}((k+1) \cdots \hat{j} \cdots n | (k+1) \cdots \hat{\imath} \cdots n), \quad (2.6)$$

which is exactly the (i, j) entry of $(A^{-1}[n-k])^{-1}$ when its rows and columns are numbered from $k+1$ to n instead of 1 to $n-k$. This proves that $A^{(k)}[n-k]$ has the form asserted. As remarked above, to do GECP is to minimize $|\det A^{(k)}[n-k]|$ at each step k, which is thus to maximize $|\det(A^{-1}[n-k])|$. □

Tornheim [1970] obtained the formula (2.6) for the diagonal entries only in $A^{(k)}[n-k]$ by direct calculation, not using Gantmacher's formula. He did not observe that this entire unreduced part of $A^{(k)}$ is actually the inverse of a matrix easily described, as we have done in (2.5).

Another idea which will be used throughout the sequel is this: if we divide A by $|a_{11}^{(0)}|$ before beginning GECP, then GECP on the new matrix will yield the same growth. So to discuss growth, we can restrict attention to CP matrices for which this has been done.

(2.7) *Definitions.* Call A *normalized* if $\max_{i,j}|a_{ij}| = 1$.* Let "CPN" abbreviate for "completely pivoted normalized." Let $\mathscr{C}_n$ denote the set of all real $n \times n$ CPN matrices. Note $a_{11} = \pm 1$ for a CPN matrix.

Clearly $g(n) = \sup\{g(A)|A \in \mathscr{C}_n\}$. The rest of this Section is devoted to proving that $g(A)$ does achieve a maximum at some B in $\mathscr{C}_n$, and that $g(n) = g(B)$ occurs at the final pivot of some such B. The subtlety involved is that, for some matrices A in $\mathscr{C}_n$, $g(A)$ will occur at an earlier diagonal entry, yet we will find $g(n)$ if we maximize $|a_{nn}^{(n-1)}|$ over $\mathscr{C}_n$. This is important because we needed a well-defined objective function for using the optimization package NPSOL to maximize $g(A)$.

The objective function used in NPSOL also must be smooth; although on the set of invertible matrices in $\mathscr{C}_n$, $a_{nn}^{(n-1)}$ is a quotient whose denominator goes to zero as A approaches a singular matrix, we will show that $|a_{nn}^{(n-1)}|$ does achieve a maximum at some B in $\mathscr{C}_n$ and is smooth on a neighborhood of B.

(2.8) *Definitions.* Let $A \in \mathscr{C}_n$. Let $h(A) = |a_{nn}^{(n-1)}|$ when A is reduced by GECP, and $h(n) = \sup\{h(A)|A \in \mathscr{C}_n\}$. Identify the set of $n \times n$ real matrices with R^{n^2} by identifying each A with the n^2-tuple obtained by concatenating the rows of A.

(2.9) PROPOSITION. *$h(A)$ is continuous on $\mathscr{C}_n$ and smooth on the invertible matrices in $\mathscr{C}_n$. $\mathscr{C}_n$ is compact, hence $h(n)$ is well defined and does achieve its sup at some invertible B in $\mathscr{C}_n$. Finally, $1 \leqslant h(n) \leqslant 2^{n-1}$ and $h(n-1) \leqslant h(n) = g(n)$. Thus $g(n)$ is well defined, does occur at some B in $\mathscr{C}_n$, and it occurs as $|b_{nn}^{(n-1)}|$.*

Proof. On R^{n^2}, $\det A$ is a polynomial function, hence is smooth. Let $A \in \mathscr{C}_n$, and reduce A by GECP. Then clearly

$$|a_{kk}^{(k-1)}| \leqslant 2|a_{k-1,k-1}^{(k-2)}| \quad \text{for each } k \geqslant 2. \tag{2.10}$$

Hence, $h(A) \leqslant 2^{n-1}|a_{11}^{(0)}| = 2^{n-1}$, the last because A is CPN. Thus $h(n)$ exists. Letting I denote the identity matrix in $\mathscr{C}_n$, $h(I) = 1$ so we have $1 \leqslant h(n) \leqslant 2^{n-1}$.

Because no row or column exchanges are done during GECP on any A in $\mathscr{C}_n$, when $\det A \neq 0$ then $a_{nn}^{(n-1)}$ is simply $(\det A)/A(1 \cdots (n-1))$. Thus on the open set of invertible matrices in $\mathscr{C}_n$, $a_{nn}^{(n-1)}$ is the quotient of nonzero smooth functions, hence itself is smooth; it is not zero, hence $h(A) = |a_{nn}^{(n-1)}|$ is also smooth on the invertible matrices in $\mathscr{C}_n$.

To show that $h(A)$ is continuous on all of $\mathscr{C}_n$, it suffices to prove that $h(A) \to 0$ if $\det A \to 0$. To do this, let $A \in \mathscr{C}_n$ and apply (2.10) to see that $|a_{kk}^{(k-1)}| \geqslant |a_{nn}^{(n-1)}|/2^{n-k}$ for any k, $1 \leqslant k < n$. Let $r = (n-1)(n-2)/2$. Then, remembering

*This use of the word "normalized" is ours and not standard; for example, the word has a different meaning in discussions of Hadamard matrices such as in Lander [1983].

that $a_{11}^{(0)} = 1$,

$$|\det A| = |a_{22}^{(1)} a_{33}^{(2)} \cdots a_{nn}^{(n-1)}| \geqslant |a_{nn}^{(n-1)}|^{n-1}/2^r.$$

Hence $h(A) = |a_{nn}^{(n-1)}|$ must go to zero if det A does.

Next observe that the set of normalized matrices in R^{n^2} is closed and bounded. Also, the set of CP matrices is closed since det A is continuous and A is CP if and only if it satisfies

$$|a_{11}| \geqslant |a_{ij}| \quad \text{for all } i, j \geqslant 1 \tag{2.11}$$

and

$$|A(1 \cdots k)| \geqslant |A(1 \cdots (k-1)i|1 \cdots (k-1)j)| \quad \text{for all } i, j \geqslant k \geqslant 2 \text{ and } k \leqslant n. \tag{2.12}$$

Since $\mathscr{C}_n$ is the intersection of these two sets, $\mathscr{C}_n$ is closed and bounded, hence compact.

To see that $h(n-1) \leqslant h(n)$, let X be an $(n-1) \times (n-1)$ matrix which is CPN and for which $h(X) = h(n-1)$. Let

$$A = \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & X \end{array}\right],$$

which is normalized because X is. Since X is CP and no actual pivoting is needed on a_{11}, A is also CP. Because $h(X) = h(n-1) \geqslant 1$, $h(A) = h(X)$. Thus $h(n) \geqslant h(A) = h(n-1)$.

Finally, we apply the existence and monotonicity of $h(n)$ to prove $g(n) = h(n)$. Let $A \in \mathscr{C}_n$; $g(A) = |a_{kk}^{(k-1)}|$ for some $k \leqslant n$, so $g(A) \leqslant h(k) \leqslant h(n)$; thus $g(A)$ is bounded above, hence $g(n)$ exists and $g(n) \leqslant h(n)$. Let $B \in \mathscr{C}_n$ such that $h(B) = h(n)$. Then $g(n) \geqslant g(B) \geqslant h(B) = h(n)$. Thus $g(n) = h(B) = h(n)$ as asserted. □

3. The best general bound known for $g(n)$. By (2.9), $g(n) \leqslant 2^{n-1}$ is true for all n. A much better bound, though still far from sharp, was proved by Wilkinson in [1961]. We will discuss that now, and how it can be mildly improved. Wilkinson obtained it by applying an old classic inequality to the determinant of each intermediate submatrix $A^{(n-k)}[k]$, followed by some clever algebraic manipulations.

Let $\alpha_n = \max\{|\det A|$: A an $n \times n$ matrix with $|a_{ij}| \leqslant 1$ for all $i, j\}$.

(3.1) HADAMARD'S INEQUALITY [Hadamard 1893]. *For each $n \geqslant 1$, $\alpha_n \leqslant n^{n/2}$, and equality holds if and only if $|a_{ij}| = 1$ for all i, j and the rows of A are mutually orthogonal.* □

It is clear geometrically that Hadamard's Inequality is true, because $|\det A|$ is the n-volume of the parallelepiped in R^n spanned by the row vectors of A (see Section IX.5, Gantmacher [1959]). Such a volume is maximized when the vectors are mutually orthogonal and every entry is ± 1, and matrices with this property have since been called *Hadamard*.

(3.2) PROPOSITION [Wilkinson 1961]. *For every $n \geqslant 1$,*

$$g(n) \leqslant \left[n\, 2\, 3^{1/2} 4^{1/3} \cdots n^{1/(n-1)}\right]^{1/2}.$$

Proof. Let A be $n \times n$ and reduce A by GECP. Let $p_1, \dots, p_n$ be the absolute values of the pivots, labeled in reverse of the ordering of rows. Thus $g(A) = p_1/p_n$, and the magnitude of every entry of $A^{(n-k)}[k]$ is bounded by p_k, so that we have:

$$|\det A^{(n-k)}[k]| = p_k p_{k-1} \cdots p_1 \leqslant p_k^k \alpha_k, \qquad k = 1, \dots, n. \tag{3.3}$$

Letting $q_k = \log(p_k)$, (3.3) implies

$$\sum_{i=1}^{n} q_i \leqslant n q_n + \log(\alpha_n) \tag{3.4}$$

and

$$\sum_{i=1}^{k-1} q_i \leqslant (k-1) q_k + \log(\alpha_k), \qquad k = 2, \dots, n-1. \tag{3.5}$$

Divide (3.4) by $n-1$ and, for each $k = 2, \dots, n-1$, divide (3.5) by $k(k-1)$. Add all these to get

$$\frac{1}{n-1} \sum_{i=1}^{n} q_i + \sum_{k=2}^{n-1} \sum_{i=1}^{k-1} \frac{q_i}{k(k-1)} \leqslant \frac{n q_n}{n-1} + \frac{\log(\alpha_n)}{n-1} + \sum_{k=2}^{n-1} \left[\frac{\log(\alpha_k)}{k(k-1)} + \frac{q_k}{k} \right]. \tag{3.6}$$

On the left side of (3.6), collect like terms and sum the telescoping series of coefficients obtained to see that the left side can be written as:

$$\frac{q_n}{(n-1)} + \left[\sum_{i=1}^{n-1} q_i/i \right].$$

Now cancel like terms in (3.6) to obtain

$$q_1 - q_n \leqslant \frac{\log(\alpha_n)}{n} + \sum_{k=2}^{n} \frac{\log(\alpha_k)}{k(k-1)}.$$

If one uses $k^{k/2}$ as a bound for every α_k, as Wilkinson did, then exponentiation now yields, as claimed,

$$(p_1/p_n) \leqslant \left[n\, 2\, 3^{1/2} 4^{1/3} \cdots n^{1/(n-1)} \right]^{1/2}. \qquad \square$$

It is not hard to get $n^{(\log n)/4}$ as an estimate for this bound (note the log of the expression is a sum which can be approximated by an integral); thus this bound is much better than 2^{n-1}. However, it is not sharp: for as Hadamard had observed, α_n can only possibly equal $n^{n/2}$ when n is $1, 2$ or a multiple of 4.

Since Hadamard's 1893 paper, somewhat smaller bounds for α_n have been obtained when n is not a multiple of 4. Brenner and Cummings [1972] give a survey of these; for example, if $n \bmod 4 = 1$, then

$$\alpha_n \leqslant \left[(n+1)^{n-1} (n-1)^{n-1} / n^{n/2} \right]^{1/2}.$$

There are similar expressions for $n \bmod 4 = 2$ or 3. If we use these bounds instead of $k^{k/2}$ for all k at the last step of the proof above, we obtain a much less elegant and only slightly smaller bound for $g(n)$. Worse, it is still far from sharp.

Indeed, the approach of the above proof cannot yield a very close bound for $g(n)$ —recall the remarks after (2.3) which pointed out that the purpose of complete

pivoting is to choose $A[k]$ within $A[k+1]$, after $n-k+1$ pivot steps, so as to minimize $|\det A^{(n-k)}[k]|$; and, for many values of k, the very structure of matrices prevents $|\det A^{(n-k)}[k]|$ from having the maximum value $p_k^k\alpha_k$.

For example, let A be an $n \times n$ Hadamard matrix; then $A^{-1} = (1/n)A^t$, so that by (2.5), $A^{(n-k)}[k] = ((1/n)A^t[k])^{-1}$. Thus to minimize the magnitude of the determinant of the latter matrix is to maximize $|\det A^t[k]| = |\det A[k]|$. However, no $k \times k$ submatrix of A can be Hadamard if $n/2 < k < n$. For let $1 \leqslant k < n$ and let B be a $k \times k$ submatrix which is Hadamard. For simplicity, suppose $B = A[k]$. Consider the $k(n-k)$-vectors which are above B in the last k columns of A. Because both A and B have mutually orthogonal columns, those k vectors must be mutually orthogonal, hence independent. Thus $k \leqslant n/2$. So if $n \geqslant 12$, there are values of k for which $n/2 < k < n$ and $k \times k$ Hadamard matrices exist, but $A[k]$ will not be such a matrix and thus $|\det A^{(n-k)}| < p_k^k\alpha_k$.

4. A new proof that $g(3) = 9/4$. We will establish the value of $g(3)$ in a more geometric way than the proofs given by Cryer [1968], Cohen [1974], and Tornheim [1964], [1970]. However, our method seems no more amenable than theirs for extension to cases of larger n.

(4.1) PROPOSITION. $g(3) = 9/4$.

Proof. We need only show $g(3) \leqslant 9/4$ since $g(A) = 9/4$ for

$$A = \begin{bmatrix} 1 & 1 & 1/2 \\ -1 & 1/2 & 1 \\ 1/2 & -1 & 1 \end{bmatrix}.$$

To begin, select any 3×3 CPN matrix A with $g(A) = g(3)$. Observe that $g(A)$ must occur in the $(3,3)$ position since $g(2)$ is clearly 2. We will make several modifications to A, if necessary, allowing us to assume it to have a rather special form that will reduce the number of cases and diagrams required for the argument. The ith row of A will be denoted by $\underline{r}_i$, and the ith row of each $A^{(k)}$ by $\underline{r}_i^{(k)}$.

First, we cannot have $\underline{r}_1 = (1,0,0)$, for if that were so then $g(3) = g(A) \leqslant 2$. Let $j = 2$ or 3 be such that $|a_{1j}| = \max\{|a_{12}|, |a_{13}|\}$. Multiply $\underline{r}_1$ by $1/a_{1j}$ and multiply the first column of A by a_{1j}. Then A will still be CPN, $g(A)$ will be unchanged, and we have that $\underline{r}_1 = (1, c, 1)$ or $\underline{r}_1 = (1, 1, c)$. Make c nonnegative by negating the second or third column of A if necessary. Now we have $0 \leqslant c \leqslant 1$ and

$$A^{(1)} = \begin{bmatrix} 1 & 1 & c \\ 0 & p & q \\ 0 & s & t \end{bmatrix} \quad \text{or} \quad A^{(1)} = \begin{bmatrix} 1 & c & 1 \\ 0 & q & p \\ 0 & t & s \end{bmatrix}.$$

Next observe that $c \neq 0$: for if $c = 0$, then $|p|, |s| \leqslant 2$ and $|q|, |t| \leqslant 1$; but then $g(A) = (|pt - qs|/\max\{|p|, |q|\}) \leqslant 2$, which is false. Thus $0 < c \leqslant 1$. Interchange columns 2 and 3 if necessary so that $\underline{r}_1$ has the form $(1, 1, c)$. Then A may no longer be CP. To accommodate this, we will allow the possibility that the second pivot element in GECP on A might be $q = a_{23}^{(1)}$ rather than $p = a_{22}^{(1)}$, in which case the second pivot step will yield $\underline{r}_3^{(2)} = (0, \pm g(A), 0)$, rather than $(0, 0, \pm g(A))$. Although this would differ from Definition (2.1), it should be clear that our approach is still general. Whether or not the difference occurs depends on the relative magnitudes of $a_{22}^{(1)}$ and $a_{23}^{(1)}$.

Negate $\underline{r}_2$ (and hence $\underline{r}_2^{(1)}$) if necessary, so that the second pivot element will be positive, whether it be $p = a_{22}^{(1)}$ or $q = a_{23}^{(1)}$.

We will use the fact that the image under a linear map of the convex hull of a finite set of points is just the convex hull of the images of those points. The rows $\underline{r}_2$ and $\underline{r}_3$ (as well as $\underline{r}_1$) lie in the convex hull of the eight points $(\pm 1, \pm 1, \pm 1)$, which we will refer to as the *cube*. When we pivot on a_{11}, the transitions $\underline{r}_i \mapsto \underline{r}_i^{(1)}$ for $i = 2, 3$ are both applications of the same linear map $(x, y, z) \mapsto (0, y - x, z - cx)$, so $\underline{r}_2^{(1)}$ and $\underline{r}_3^{(1)}$ both lie in the convex hull of the images under this map of the points $(\pm 1, \pm 1, \pm 1)$.

Henceforth we identify any point $(0, y, z)$ in R^3 with (y, z) in R^2. Since $0 < c \leqslant 1$, we see that $\underline{r}_2^{(1)}$ and $\underline{r}_3^{(1)}$ lie in the convex hull H of the six points $(2, -1 + c)$, $(2, 1 + c)$, $(0, 1 + c)$, $(-2, 1 - c)$, $(-2, -1 - c)$, and $(0, -1 - c)$, which is the following hexagon.

We now make the final modification to A, altering $\underline{r}_2$ if necessary so that $\underline{r}_2^{(1)}$ lies on the boundary of H. To see that this is possible, since it is clear that $\underline{r}_2^{(1)} \neq (0, 0)$, if $\underline{r}_2^{(1)}$ lies in the interior of H we can find $\lambda > 1$ such that $\lambda \underline{r}_2^{(1)}$ lies on the boundary of H. Replace $\underline{r}_2$ by a preimage in the cube of $\lambda \underline{r}_2^{(1)}$ under the above linear map. Then $\underline{r}_2^{(1)}$ will lie on the boundary of H, and all the properties we have already assumed for A will still hold.

By our assumptions, $\underline{r}_2^{(1)} = (p, q)$ lies on one of the line segments numbered from 1 to 7 in FIGURE (4.1). For each of these seven segments, if (p, q) lies on it we can determine the largest possible value for $h(A)$ in terms of p, q, and c. We will then maximize this over p and q, and then finally over c.

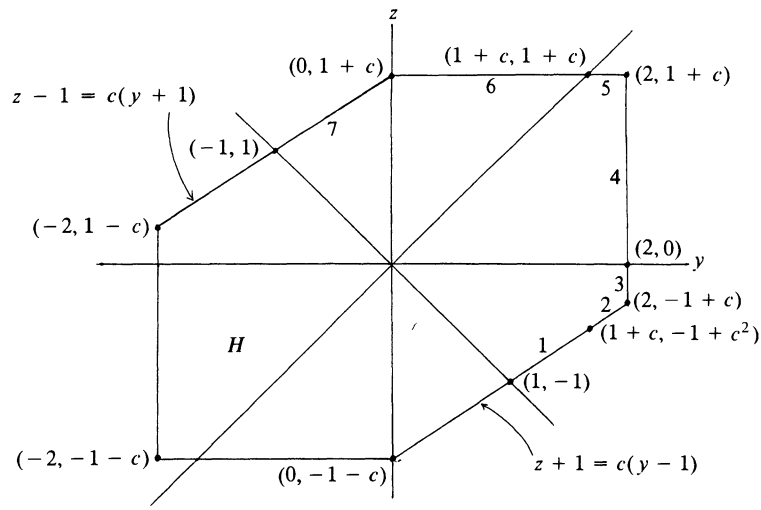


FIG. (4.1).

In more detail, assume (p, q) is a particular point on line segment number i. If $1 \leqslant i \leqslant 5$ then $p = a_{22}^{(1)}$ will be the second pivot element in GECP, but if $6 \leqslant i \leqslant 7$ then $q = a_{23}^{(1)}$ will be the second pivot element. In either case let $\mu = \max\{|p|, |q|\} = \max\{p, q\}$ and let $S = S_{p,q} = \{(y, z) \in R^3 |\ |y|, |z| \leqslant \mu\}$. Then $\underline{r}_3^{(1)} = (s, t)$

must lie in $H \cap S$. Now $H \cap S$ is itself the convex hull of its finite set of extreme points, which may be given in terms of p, q, and c. This is illustrated in FIGURE (4.2) for $i = 1$, and tabulated for all $1 \leqslant i \leqslant 7$ in the first column of TABLE (4.3).

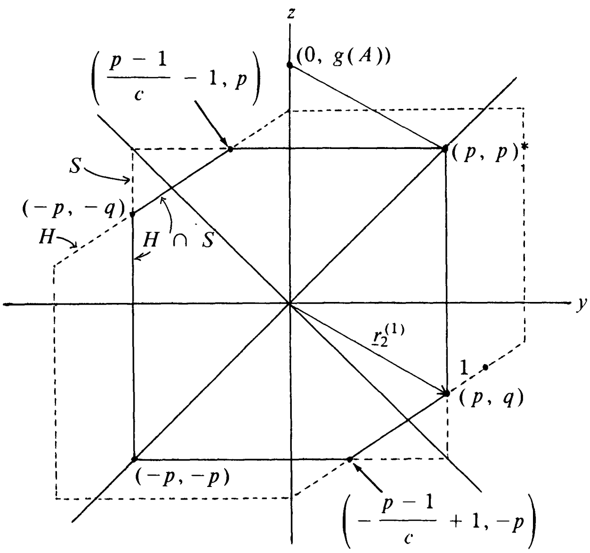


FIG. (4.2)

The second pivot step, which transforms $r_3^{(1)}$ into $r_3^{(2)}$, is again a linear map, either $(s, t) \mapsto (0, t - sq/p)$ or $(s, t) \mapsto (s - tp/q, 0)$, depending respectively on whether $1 \leqslant i \leqslant 5$ or $6 \leqslant i \leqslant 7$. So $r_3^{(2)}$ lies in the convex hull of the images under this linear map of the extreme points of $H \cap S$. For each value of i, apply this linear map to each extreme point of $H \cap S$, and so determine the largest possible value of $h(A)$ in terms of p, q, and c. The result is given in the third column of TABLE (4.3), and in the second column we indicate by an * one of the extreme points of $H \cap S$ which makes this largest possible value of $h(A)$ occur. Since each result is linear in p and q, it is itself maximized when (p, q) is one of the endpoints of the ith interval. This endpoint is listed in the fourth column of TABLE (4.3), and the fifth column gives the corresponding maximum possible value of $h(A)$ in terms of c alone. In every case, we see that $h(A) \leqslant 2 + c(1 - c)$. Finally, since $2 + c(1 - c) \leqslant 9/4$ for all real c, we have shown that $h(A) \leqslant 9/4$. Thus $g(3) = h(A) \leqslant 9/4$ also, and the proof is complete. □

5. The growth conjecture for Hadamard matrices. As discussed in Section 3, Hadamard showed in [1893] that the maximum possible magnitude for the determinant of a normalized $n \times n$ matrix A is $n^{n/2}$, and this maximum occurs if and only if A has mutually orthogonal rows and each entry ± 1. Such matrices have since been called *Hadamard*. Hadamard constructed some examples; in particular, he observed as Sylvester had done in [1867] that such matrices do exist when n is a

TABLE (4.3)

Segment containing $r_2^{(1)} = (p, q)$	Extreme points of $H \cap S$	Largest possible $h(A)$ for fixed p, q, c	Optimal (p, q)	$g(A)$ at optimal (p, q)
1	$(p, q), (p, p)^*, \left(\frac{p-1}{c} - 1, p\right)$ $(-p, -q), (-p, -p),$ $\left(-\frac{p-1}{c} + 1, -p\right)$	$p - q$	$(1 + c, -1 + c^2)$	$2 + c(1 - c)$
2	$(p, q), (p, 1 + c)^*, (0, 1 + c)$ $(-p, -q), (-p, -1 - c),$ $(0, -1 - c)$	$1 + c - q$	$(1 + c, -1 + c^2)$	$2 + c(1 - c)$
3	$(2, -1 + c), (2, 1 + c)^*,$ $(0, 1 + c), (-2, 1 - c),$ $(-2, -1 - c), (0, -1 - c)$	$1 + c - q$	$(2, -1 + c)$	2
4	$(2, -1 + a), (2, 1 + c),$ $(0, 1 + c)^*, (-2, 1 - c),$ $(-2, -1 - c), (0, -1 - c)$	$1 + c$	anywhere	$1 + c$
5	$(p, 1 + c), (0, 1 + c),$ $(-p, 1 - c(p - 1))^*, (-p, -1 - c),$ $(0, -1 - c), (p, -1 + c(p - 1))$	$2 + c - c(p - 1)$	$(1 + c, 1 + c)$	$2 + c(1 - c)$
6	$(1 + c, 1 + c), (0, 1 + c),$ $(-1 - c, 1 - c^2), (-1 - c, -1 - c),$ $(0, -1 - c), (1 + c, -1 + c^2)^*$	$(1 + c) +$ $(1 - c)p$	$(1 + c, 1 + c)$	$2 + c(1 - c)$
7	$(p, q), (-q, 1 - c(q - 1)),$ $(-q, -q), (-p, -q),$ $(q, -1 + c(q - 1)), (q, q)^*$	$q - p$	$(0, 1 + c)$	$1 + c$

power of 2. The simplest examples of Hadamard matrices are in these sizes:

(5.1) *Definition.*

$$H_1 = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}, \; H_2 = \begin{bmatrix} H_1 & H_1 \\ -H_1 & H_1 \end{bmatrix}, \quad \text{and in general,} \quad H_{k+1} = \begin{bmatrix} H_k & H_k \\ -H_k & H_k \end{bmatrix}.$$

Hadamard showed that, if an $n \times n$ Hadamard matrix exists for $n > 2$ then n is a multiple of 4. He constructed 12×12 and 20×20 examples, and asked what are all the values of n for which they exist. It seems generally believed that they do exist for every multiple of 4, but that is not known yet; the earliest reference in which we found this conjecture is Paley [1933]. The author remarked prophetically that the conjecture "seems probable... but the general theorem has every appearance of difficulty." There is a large and growing body of literature on existence and applications of Hadamard matrices, and they have been constructed for many infinite families of values of n (see for example Lander [1983] and Hall [1986]).

Hadamard matrices are of interest here because all the evidence indicates that they give maximum possible growth in the dimensions where they exist, and no other type matrix does. It seems reasonable that, with their many strong properties,

it should be possible to prove Wilkinson's conjecture for them more easily than for general matrices. We have not done that, but our results below strengthen the evidence that the conjecture is true for them.

We should emphasize that, even if one knew what growth is for ± 1 matrices, it is not clear how that would help determine $g(n)$ in general: it is easily seen that $g(3) = 2.25$ does not occur in such a matrix. On the other hand, if one could prove that $g(n)$ is achieved on a Hadamard matrix in dimensions n where one exists, that would be valuable: since they definitely exist for n any power of 2, it would follow easily from the monotonicity of $g(n)$ that $g(n) < 2n$ for all n. That is not as good as Wilkinson's conjectured bound n, but much better than the bound given in Section 3.

When Gaussian elimination is done on an $n \times n$ Hadamard matrix A, the last pivot has magnitude n. This was observed by Tornheim [1964] and Cryer [1968], because it is the reciprocal of an entry from A^{-1} and that equals $(1/n)A^t$. Thus $g(A) \geqslant n$. Cryer also evaluated the two pivots preceding the last by invoking results of Sharpe [1907] which show that the nonzero $(n-2) \times (n-2)$ subdeterminants of an $n \times n$ Hadamard matrix all have the same magnitude, as do the nonzero $(n-3) \times (n-3)$ subdeterminants. Cryer remarked it is unlikely that any earlier pivot under GECP could exceed n, hence it seems very probable that $g(A) = n$.

Our Proposition (5.5) below provides a new proof and extension of these results of Sharpe, Tornheim and Cryer, and strengthens the plausibility of Cryer's remark. The key tool is Gantmacher's formula, stated in Proposition (2.4) above. As shown next, when A^{-1} is proportional to A^t, in particular when A is Hadamard, that formula provides a proportionality between any subdeterminant of A and its complementary subdeterminant in A. Thus the possible values of the largest subdeterminants of a Hadamard matrix can be evaluated in terms of small ones, whose values are easy to list.

(5.2) PROPOSITION. *Let A be an invertible $n \times n$ matrix for which $AA^t = cI$. Then $A^{-1} = (1/c)A^t$ and $\det A = c^{n/2}$. If $1 \leqslant k < n$, then*

$$A(12\cdots k) = \pm c^{k-n/2}A((k+1)\cdots n) = \pm c^{k-n/2}\det A[n-k].$$

Proof. The $(1,1)$ entry of AA^t is the square of the Euclidean length of row 1 of A, hence $c > 0$. Since $AA^t = cI$, $A^{-1} = (1/c)A^t$ and also $(\det A)^2 = c^n$, hence $\det A = \pm c^{n/2}$. Let $X = (1/c)A^t$, so that

$$X(12\cdots k) = c^{-k}A(12\cdots k)$$

and by (2.4) we also have

$$X(12\cdots k) = \pm c^{-n/2}A((k+1)\cdots n).$$

Two more preliminary results will be useful. The next one includes a special interpretation of what Wilkinson's conjecture means for Hadamard matrices.

(5.3) PROPOSITION. *Let A be an $n \times n$ Hadamard matrix and reduce A by Gaussian Elimination. No matter how pivots are chosen, after k pivot steps, the unreduced submatrix of $A^{(k)}$ satisfies*

$$A^{(k)}[n-k] = n\left(A^t[n-k]\right)^{-1}.$$

Thus, when GECP is done, Wilkinson's conjecture that $g(A) \leqslant n$ is true if and only if,

for every $p \times p$ lower right submatrix $A[p]$, the magnitude of every entry of $(A[p])^{-1}$ is less than or equal to 1, $p = 1, \dots, n$.

Proof. Since $A^{-1} = (1/n)A^t$, $A^{-1}[n-k] = (1/n)A^t[n-k]$; this with (2.5) implies $A^{(k)}[n-k] = (A^{-1}[n-k])^{-1} = n(A^t[n-k])^{-1}$.

We need the following facts. The first two are simple and known; see, for example, Brenner and Cummings [1972] or MacWilliams and Sloane [1977].

(5.4) PROPOSITION. *Let B be $m \times m$ and $b_{ij} = \pm 1$ for all i, j. Then*
(i) *$\det B$ is an integer and 2^{m-1} divides $\det B$.*
(ii) *When $m \leqslant 6$, the only possible values for $|\det B|$ are these, and they do all occur*:

m	1	2	3	4	5	6
$\lvert\det B\rvert$	1	0, 2	0, 4	0, 8, 16	0, 16, 32, 48	0, 32, 64, 96, 128, 160

(iii) *Suppose that $\det B \neq 0$ and every $(m-1) \times (m-1)$ subdeterminant of B is zero or has the minimum possible nonzero magnitude, 2^{m-2}. Then every nonzero $(m-2) \times (m-2)$ subdeterminant of B has magnitude 2^{m-3} or 2^{m-2}.*

Proof. To prove (i) and (ii), do one step of Gaussian Elimination on B. The new matrix $B^{(1)}$ has only ± 1 in row 1 and only 0 or ± 2 in rows 2 through m. Thus $\det B = 2^{m-1}(\det C)$, and C is a matrix having only 0 and ± 1 entries so $\det C$ is an integer. The values listed for $|\det B|$ in each category $m \leqslant 5$ above are the multiples of 2^{m-1} less than or equal to $m^{m/2}$, that being an upper bound by Hadamard's Inequality (3.1). For $m = 6$, every such value except 192 is listed, which is known not to occur (see Brenner and Cummings [1972]). It is not difficult to construct examples of each size listed.

For (iii), suppose that $\det B \neq 0$ and every nonzero $(m-1) \times (m-1)$ subdeterminant of B has magnitude 2^{m-2}. Then $B^{-1} = (2^{m-2}/(\det B))C$, where every entry of C is 0 or ± 1. Choose any i_1, i_2. Then clearly $|C(i_1 i_2)| = 0, 1$ or 2 so

$$|B^{-1}(i_1 i_2)| = \left(2^{m-2}/(\det B)\right)^2 d, \qquad d = 0, 1 \text{ or } 2.$$

By (2.4), we also have

$$B^{-1}(i_1 i_2) = \left(1/(\det B)\right) B(i_3' \cdots i_m').$$

So $B(i_3' \cdots i_m')$ is an arbitrary $(m-2) \times (m-2)$ subdeterminant of B, and if it is not 0, then

$$|\det B| = \frac{2^{2m-4}}{B(i_3' \cdots i_m')} d, \qquad d = 1 \text{ or } 2.$$

Since $\det B$ must be a multiple of 2^{m-1}, and $B(i_3' \cdots i_m')$ is a multiple of 2^{m-3}, it follows that $|B(i_3' \cdots i_m')|$ can only equal 2^{m-3} or 2^{m-2}.

The next result provides a new proof and extends to more cases the results of Sharpe [1907] and Cryer [1986] discussed above.

(5.5) PROPOSITION. *Let A be an $n \times n$ Hadamard, and reduce A by Gaussian Elimination, not necessarily with complete pivoting. Let $D = n^{n/2}$. For simplicity of notation, suppose no row or column exchanges are needed. Then the possible values for*

$|A(1 \cdots k)|$ and $|a_{kk}^{(k-1)}|$ when $k \geqslant n-6$ are:

TABLE (5.6)

k	$\lvert A(1 \cdots k)\rvert$	$\lvert a_{kk}^{(k-1)}\rvert$
n	D	n
$n-1$	$(1/n)D$	$n/2$
$n-2$	$(2/n^2)D$	$n/2$
$n-3$	$(4/n^3)D$	$n/2$ or $n/4$
$n-4$	$(q/n^4D, q = 8, 16$	$np, p = 1, 1/2, 1/4, 1/3, 1/6$
$n-5$	$(q/n^5)D, q = 16, 32, 48$	$np, p = 3/2, 1, 3/4, 1/2, 3/8, 1/3, 3/10, 1/4, 1/5, 1/8, 1/10$
$n-6$	$(q/n^6)D, q = 32, 64, 96, 128, 160$	

In fact, the above values for $|A(1 \cdots k)|$ are all possible values for the magnitude of any nonzero $k \times k$ subdeterminant of A, $k = n-6, \ldots, n$.

If complete pivoting is done, then $|a_{n-5,n-5}^{(n-4)}|$ cannot be $(3/2)n$ and thus the magnitudes of the six last pivots for GECP are at most n, n, $n/2$, $n/2$, $n/2$, and n, respectively.

Proof. Note that $A(1 \cdots k) \neq 0$ since Gaussian Elimination can be completed without row or column exchanges. Then $\det A[n-k] = A((k+1) \cdots n)$ cannot be zero for any k because A is Hadamard, by (5.2). We will use these facts without further comment.

Since $|A(1 \cdots k)| = n^{k-n/2}|\det A[n-k]| = Dn^{k-n}|\det A[n-k]|$, the values for determinants of $6 \times 6 \pm 1$ matrices given in Proposition (5.3) imply that $|A(1 \cdots (n-6))| = Dn^{-6}p$, where $p = 32, 64, 96, 128$, or 160. In the same way, we can obtain all the possible values for $|A(1 \cdots k)|$, $k = n-5, \cdots, n$, and these are all listed in Table (5.6).

Now $a_{kk}^{(k-1)} = A(1 \cdots k)/A(1 \cdots (k-1))$ so the only possible values for that are the quotients obtained from the previous lists; for example, $|a_{n-5,n-5}^{(k-6)}|$ could only equal nq/p where $q = 16, 32$ or 48 and $p = 32, 64, \ldots, 160$. In the same way we can calculate all possible values for $|a_{kk}^{(k-1)}|$, $k = n-4, \ldots, n$. All such quotients are listed in TABLE (5.6). The above values apply to any nonzero $k \times k$ subdeterminant of A, because any one could be put into the leading principal $k \times k$ position originally; while some exchanges might be needed within the first k rows during Gaussian Elimination, this would not change the value of this subdeterminant.

Suppose now that complete pivoting is done, i.e., that A is CP. We will show that $|a_{n-5,n-5}^{(n-4)}|$ cannot equal $n(48/32)$, which is the only one of all the possible values for the last six pivots which would violate Wilkinson's conjecture.

By way of contradiction, suppose that $|\det A[6]| = 32$ and $|\det A[5]| = 48$. There is actually a $6 \times 6 \pm 1$ matrix like that, but we will see that it cannot be the lower right principal 6×6 submatrix of a CP Hadamard matrix.

Consider $A[7]$, the lower right 7×7 principal submatrix of A. We claim that A being CP and $|\det A[6]| = 32$ implies that every nonzero 6×6 subdeterminant of $A[7]$ has the minimum possible magnitude 32. For if any one were larger, we could rearrange only the last seven rows and columns of A so the new lower right principal 6×6 subdeterminant of A is larger than 32; then the complementary subdeterminant of this in A is larger than the one originally in that position, by (5.2). Thus the quotient of subdeterminants of A which yields $a_{n-5,n-5}^{(n-6)}$ is larger

than the original such quotient, which is contradictory since A was assumed to be CP. This allows us to apply (5.4) (iii) to conclude that each nonzero 5×5 subdeterminant of $A[7]$ has magnitude 16 or 32. Thus $|\det A[5]|$ cannot be 48.

We believe that more of the values listed above for pivots are impossible when GECP is done on a Hadamard. In particular, we conjecture that every pivot before $|a_{nn}^{(n-1)}|$ will have magnitude at most $n/2$. We also believe that the fourth from the last pivot can only have magnitude $n/4$, not $n/2$. The latter conjecture developed as we experimented to create different pivot patterns with Hadamard matrices. We always saw $\pm n/4$ as the fourth from last pivot, even when we tried to force $\pm n/2$ there. Also, we suspect that there is some connection between the possible values for the kth and kth from last pivots when GECP is done; if so, since the fourth pivot has to be ± 4 by Proposition (5.8) below, it does not seem unreasonable that the fourth from last should also have a unique possible magnitude.

The next well known result is usually proved by describing the sign patterns in rows, much as we do in the first part of the proof of (5.8) below. However, it is an easy consequence of Propositions (5.4) and (3.1).

(5.7) PROPOSITION (Hadamard [1893]). *If A is an $n \times n$ Hadamard matrix and $n > 2$, then n is a multiple of* 4.

Proof. Since $n > 2$, n is even by (5.4). Let $n = 2k$, $k \geqslant 1$. By (3.1), $\det A = (2k)^k$ and this is divisible by 2^{2k-1} by (5.4) (i). The quotient is $k^k/2^{(k-1)}$, which can be integral only if k is even.

(5.8) PROPOSITION. *Let $n \geqslant 4$ and A be an $n \times n$ Hadamard. Reduce A by GECP. Then the magnitudes of the first four pivots are* $1, 2, 2$ *and* 4; *and if $n > 4$, $|a_{55}^{(4)}|$ is* 2 *or* 3.

Proof. We may suppose that A is CP. Clearly $|a_{11}^{(0)}| = 1$, and every entry of $A^{(1)}$ is 0 or ± 2. Thus $a_{22}^{(1)} = \pm 2$ and every entry of $A^{(2)}$ must be 0, ± 2 or ± 4. Because $g(3) < 4$, there cannot be any 4's in $A^{(2)}$. Hence every entry of $A^{(2)}$ is 0 or ± 2 so $a_{33}^{(2)}$ must be ± 2, and then every entry of $A^{(3)}$ is 0, ± 2 or ± 4.

We will now show that there is an entry ± 4 in the last $n-3$ rows and columns of $A^{(3)}$, hence $|a_{44}^{(3)}|$ has to be 4 because A is completely pivoted.

By (5.7), $n = 4p$ for some $p \geqslant 1$. Without destroying the CP property of A, we can change signs of rows and columns to make the first row and column positive. Then each column of A has one of the following four sign patterns in its first three entries:

$$\begin{array}{cccc} \mathrm{I} & \mathrm{II} & \mathrm{III} & \mathrm{IV} \\ + & + & + & + \\ + & - & + & - \\ + & + & - & - \end{array} \tag{5.9}$$

Further, the mutual orthogonality of the first three rows of A implies that there are exactly p columns of each type. Since $A(123)$ is not zero, the first three columns of A must be of three different column types.

Choose any column j of the fourth type not represented among columns 1, 2 and 3. There are p of these, so at least one exists. For the purpose of finding a row i so that $A(123i|123j) = \pm 16$, it does not matter if columns 2, 3 and j are rearranged; thus we may assume that A has the pattern (5.9) in its first three rows and the four columns $1, 2, 3, j$.

Rows can be divided into the same four types of groups of p each as columns can, so there are p rows having pattern $+--$ in the first three entries, and all these lie below the first three rows. At least one of these rows must have $a_{ij}=+1$: for otherwise column j of A has more than p negative entries, which contradicts its being orthogonal to column 1. Thus $A(123i|123j)$ is the determinant of the following matrix, and that is 16:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

Then $a_{ij}^{(3)} = A(123j|123i)/A(123) = \pm 16/4 = \pm 4$ as desired.

Now let $n > 4$ and let us examine the fifth pivot. Because every entry in $A^{(3)}$ is of magnitude 0, 2 or 4, pivoting on $a_{44}^{(3)}$ will only involve adding ± 1 or $\pm(1/2)$ times the fourth row of $A^{(3)}$ to the rows below, and this will create only integer entries in $A^{(4)}$. Thus $|a_{55}^{(4)}|$ must be 1, 2 or 3 in order to be an integer and for $|A(12345)| = 16|a_{55}^{(4)}|$ to be less than $5^{5/2}$. It will be 2 for the 8×8 Hadamard H_3 defined in (5.1); and it will be 3 if GECP is done on a 12×12 Hadamard such as the one in [Paley 1933]. To see that it cannot be 1 is to show that one could not have

$$A^{(4)} = \left[\begin{array}{cccc|c} 1\ldots\ldots & & & & \ldots\ldots \\ & 2\ldots & & & \ldots\ldots \\ & & 2\ldots & & \ldots\ldots \\ & & & 4\ldots & \ldots\ldots \\ \hline & 0 & & & B \end{array}\right],$$

where every entry of B is zero or ± 1: for if that were true, then B would be a normalized $(n-4)\times(n-4)$ matrix, so $|\det B| \leqslant (n-4)^{(n-4)/2}$ by (3.1). But $|\det B| = (n^{n/2})/16$, and it is easily checked that these cannot both hold when $n > 4$.

We will now show that Wilkinson's conjecture does hold for one class of Hadamard matrices, those defined in (5.1). Tornheim proved this in [1970] and we rediscovered it. Proposition (5.12) is a mild generalization of his theorem and stated in our notation. A definition is needed first.

(5.11) *Definition.* Let A be $n\times p$ and let B be any rectangular matrix. The *Kronecker product* of A and B is

$$A\otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \ldots & a_{1p}B \\ a_{21}B & a_{22}B & \ldots & a_{2p}B \\ \cdot & \cdot & \ldots & \cdot \\ \vdots & & \cdot & \vdots \\ a_{n1}B & a_{n2}B & \ldots & a_{np}B \end{bmatrix}.$$

Let $\otimes^k A$ denote the Kronecker product of k copies of A, which is well defined since the Kronecker product is associative (see Lancaster and Tismenetsky [1985]). Thus the Hadamard matrix H_n defined in (5.1) is $\otimes^n H_1$.

(5.12) PROPOSITION [Tornheim 1970]. *Suppose A is $n\times n$ and CP. Then $A\otimes H_1$ is CP and its pivots are the Kronecker product of the pivots of A with those of H_1—that*

is, the entries of

$$\left[a_{11}^{(0)} \cdots a_{nn}^{(n-1)}\right] \otimes [1\,2].$$

Therefore, $A \otimes H_n$ is CP for all n and the pivots are the Kronecker product of the pivots of A with $\otimes^n [1\,2] = [1\,2\,2\,4\,2\,4\,4\,8\ldots]$.

Proof. Suppose that two steps of GECP are done on

$$A \otimes H_1 = \begin{bmatrix} a_{11}H_1 & a_{12}H_1 & \ldots & a_{1n}H_1 \\ \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & & \cdot \\ a_{n1}H_1 & a_{n2}H_1 & \ldots & a_{nn}H_1 \end{bmatrix}.$$

If no row or column exchanges have been done, then for $i > 1$, the block $a_{ij}H_1$ must have been changed into $a_{ij}H_1 - (a_{1j}a_{j1}/a_{11})H_1$ and the block $a_{1j}H_1$ into $a_{1j} \otimes H_1^{(1)}$. Thus

$$(A \otimes H_1)^{(2)} = \left[\begin{array}{c|c} a_{11} \otimes H_1^{(1)} & \cdots\ a_{1n} \otimes H_1^{(1)} \\ \hline 0 & A^{(1)}[n-1] \otimes H_1 \end{array}\right]. \tag{5.13}$$

To see that no exchanges will be done in those first two steps, observe that the entries of $A \otimes H_1$ are $\pm a_{ij}$, hence $|a_{11}|$ dominates each because A is CP. So the $(1,1)$ entry will be the first pivot. After the first pivot step the upper left 2×2 block and the block $a_{ij}H_1$ for $i > 1$ have become, respectively,

$$\begin{bmatrix} a_{11} & a_{11} \\ 0 & 2a_{11} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} a_{ij}^{(1)} & a_{ij}^{(1)} \\ -a_{ij}^{(1)} & a_{ij} + (a_{1j}a_{j1})/a_{11} \end{bmatrix}.$$

Since A is CP, each entry of the second type block is dominated by $|2a_{11}|$, so no exchange is done at the second step of GECP either. Therefore, two steps of GECP do produce $(A \otimes h_1)^{(2)}$ as shown in (5.12). The rest of the proof that $A \otimes H_1$ is CP now follows by induction. The final upper triangular matrix is $(A \otimes H_1)^{(2n-1)} = A^{(n-1)} \otimes H_1^{(1)}$, hence the pivots are as claimed. For $n > 1$, $A \otimes H_n$ equals $(A \otimes H_{n-1}) \otimes H_1$ so by induction, $A \otimes H_n$ is CP for all n; and finally, H_1 is CP and $H_n = H_1 \otimes H_{n-1}$, hence H_n is CP for all $n \geqslant 1$.

Curiously enough, it is false that $H_1 \otimes A$ need be CP if A is—just consider $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. This example also shows of course that having A and B completely pivoted is not enough to imply that $A \otimes B$ is also.

The above results show that when GECP is done on Hadamard matrices, the magnitudes of a few of the first and last pivot elements are determined. However, we obtained examples which show that the entire sequence of magnitudes of pivot elements in the row reduction of an $n \times n$ CP Hadamard matrix is not determined by n, nor even by the Hadamard class of that matrix, when $n > 12$. This says that there is a rich variety of values among the various intermediate size subdeterminants in such a matrix.

(5.14) *Definition.* Two matrices are called *Hadamard equivalent* if one can be obtained from the other by a sequence of row exchanges, column exchanges, row negations, and column negations.

H_2, and H_3 are representatives for $n = 2, 4$, and 8; and an example in Paley [1933] is a representative for $n = 12$. For $n = 16$ it is proved in Hall [1961] that there are 5 classes, and examples of each are given. For $n = 20$ there are 3 classes, with proof and examples in Hall [1965]. Also see Hall [1986].

When an $n \times n$ Hadamard matrix A is reduced by GECP, let $\underline{p} = (p_1, \ldots, p_n)$ where $p_k = |a_{kk}^{(k-1)}|$. (Note this is the reverse of how p_i was defined in Section 3.) If $n = 2$, 4, or 8, then necessarily $\underline{p} = (1,2)$, $\underline{p} = (1,2,2,4)$, or $\underline{p} = (1,2,2,4,2,4,4,8)$ by the results in (5.5) and (5.8) above. For $n = 12$, we performed Gaussian Elimination on many different 12×12 CP Hadamard matrices (all Hadamard equivalent to one another). We always obtained $\underline{p} = (1,2,2,4,3,10/3,18/5, 4,3,6,6,12)$, and conjecture that this is the only pivot pattern possible when $n = 12$.

For 16×16 Hadamard matrices, in Hall [1961] the 5 classes are numbered I, ..., V. Classes IV and V are one another's transpose, and so are identical for the purpose of GECP (since A is CP if and only if A^t is, by the discussion following Proposition (2.3), in which case both give the same $\underline{p}$). Within each of classes I–IV we row reduced many CP representatives and obtained various results for $\underline{p}$, as indicated below.

TABLE (5.15)

Pivot Patterns for 16×16 *Hadamard Matrices Reduced by GECP*

Class	$\underline{p}$
	$(1,2,2,4,2,4,\ 4,8,2,\ 4,4,8,\ 4,8,8,16)$
I	$(1,2,2,4,3,8/3,\ 4,6,8/3,4,4,8,\ 4,8,8,16)$
	$(1,2,2,4,3,8/3,\ 4,6,8/3,4,6,16/3,\ 4,8,8,16)$
	$(1,2,2,4,2,4,\ 4,\ 8,2,\ 4,4,8,\ 4,8,8,16)$
II	$(1,2,2,4,2,4,\ 4,\ 6,8/3,\ 4,6,16/3,\ 4,8,8,16)$
	$(1,2,2,4,3,10/3,\ 12/5,\ 4,16/3,\ 4,4,8,\ 4,8,8,16)$
	$(1,2,2,4,2,4,\ 4,\ 8,2,\ 4,\ 4,\ 8,\ 4,8,8,16)$
III	$(1,2,2,4,2,4,\ 4,\ 6,8/3,\ 4,\ 4,\ 8,\ 4,8,8,16)$
	$(1,2,2,4,3,10/3,\ 18/5,\ 4,4,\ 40/9,\ 24/5,\ 16/3,\ 4,8,8,16)$
	$(1,2,2,4,2,4,\ 4,\ 4,4,4,\ 4,\ 8,\ 4,8,8,16)$
IV	$(1,2,2,4,2,4,\ 4,\ 4,4,5,\ 24/5,\ 16/3,\ 4,8,8,16)$
	$(1,2,2,4,3,10/3,\ 18/5,\ 4,4,40/9,\ 24/5,\ 16/3,\ 4,8,8,16)$

Observe that $p_{n-3} = n/4$ in every case, as mentioned after Proposition (5.5) above.

6. NPSOL and the growth and length problems. As observed earlier, examples of CPN $n \times n$ matrices A for $2 \leqslant n \leqslant 4$ which satisfy $g(A) = g(n)$ are given by

$$A = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 1 & 1/2 \\ -1 & 1/2 & 1 \\ 1/2 & -1 & 1 \end{bmatrix}, \text{ and } A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

Each of these also satisfies $AA^t = cI$ and $g(A) = c$. This observation led us to conjecture that the same phenomenon might hold for larger n. We now know that this conjecture is almost certainly false for $n = 5$ and $n = 7$, as we will explain below. However, in early attempts to gather evidence in favor of the conjecture, and to generate examples of matrices with large growth, we were led to obtain the

non-linear programming software package "NPSOL," written by Gill, Murray, Saunders, and Wright of the Systems Optimization Laboratory group in the Operations Research Department at Stanford University [NPSOL 1983]. A detailed discussion of the algorithm employed in NPSOL (sequential quadratic programming) can be found in Gill, Murray, and Wright [1981]. We will describe what this package does and how we used it, and then discuss the rather intriguing results we obtained and questions that these results suggest. The following definitions will be useful; they are ours, not standard ones.

(6.1) *Definition.* We call a normalized $n \times n$ matrix A *normalized orthogonal* (NO) if $AA^t = c(A)I$ for some constant $c(A)$. Let $c(n) = \sup\{c(A)|A \text{ an } n \times n \text{ NO matrix}\}$.

If we identify the set of $n \times n$ real matrices with R^{n^2} as in Section 2, it is clear that the NO matrices form a compact subset of R^{n^2} on which the function $c(A)$ is continuous. Thus $c(n)$ exists and there is an $n \times n$ NO matrix A with $c(A) = c(n)$. Also, it is clear that $c(n) \leqslant n$, and $c(n) = n$ iff there exists an $n \times n$ Hadamard matrix. So the determination of $c(n)$, which we call the *length problem* because $\sqrt{c(A)}$ is the usual Euclidean length of each row of an NO matrix A, is at least as difficult as the determination of those n for which an $n \times n$ Hadamard matrix exists.

For any $n \times n$ NO matrix A, $A^{-1} = (1/c(A))A^t$; hence by Proposition (2.3) we have $h(A) = c(A)/|a_{ij}|$ where a_{ij} is some non-zero entry of A. Since $g(n) = h(n)$ by Proposition (2.9), it follows that $g(n) \geqslant c(n)$. If it were true that $g(n) = c(n)$ for all n, Wilkinson's conjecture would follow at once. However, as mentioned above and explained below, this is almost certainly false for $n = 5$ and $n = 7$.

The NPSOL package addresses the problem of finding $\min f(\underline{x})$ where $\underline{x} = (x_1, \ldots, x_m) \in R^m$ is subject to the constraints

$$k_j \leqslant x_j \leqslant u_j \quad \text{for} \quad 1 \leqslant j \leqslant m \quad \text{and}$$

$$k_{m+j} \leqslant g_j(\underline{x}) \leqslant u_{m+j} \quad \text{for} \quad 1 \leqslant j \leqslant p.$$

The functions $f, g_1, \ldots, g_p$ must be continuously differentiable and the bounds k_j and u_j may be equal, or may be infinite. The user of the package must write a Fortran program to call the main subroutine of the package, and must also write subroutines to be called from within the package to evaluate $f, g_1, \ldots, g_p$ and all of their first partial derivatives at a specific point $\underline{x}$. An initial point $\underline{x}$ must be supplied which satisfies all the constraints, and the package moves $\underline{x}$ to a local constrained minimum for f. Of course if one wants to maximize f, one minimizes $-f$. The package returns an integer variable "INFORM" which, if equal to 0, indicates that the authors of the package consider the result returned to be a reliable local minimum. We will refer to a run as *successful* if it returns with INFORM = 0.

For the length problem, that is, to determine $c(n)$, we used NPSOL in the following way. Let x_{ij} be the n^2 variables ($1 \leqslant i, j \leqslant n$) representing the entries of an $n \times n$ matrix, with variable bounds $-1 \leqslant x_{ij} \leqslant 1$. The objective function is $f(\underline{x}) = -\sum_{j=1}^{n} x_{1j}^2$, and the constraints are given by

$$0 \leqslant \sum_{j=1}^{n} x_{ij}^2 - \sum_{j=1}^{n} x_{i+1,j}^2 \leqslant 0 \qquad (1 \leqslant i \leqslant n-1)$$

and

$$0 \leqslant \sum_{k=1}^{n} x_{ik}x_{jk} \leqslant 0 \qquad (1 \leqslant i < j \leqslant n).$$

It was necessary to reformulate all of these in terms of variables $x_1, \dots, x_{n^2}$. Analytic formulae for all the relevant partial derivatives are easily obtained, and writing the required subroutines was tedious but straightforward.

For the growth problem, that is to find $h(n) = g(n)$, again we let x_{ij} $(1 \leqslant i, j \leqslant n)$ represent entries of an $n \times n$ matrix $\underset{\sim}{x}$, with variable bounds $-1 \leqslant x_{ij} \leqslant 1$. We required $x_{11} = 1$. The objective function is $-(x_{nn}^{(n-1)})^2$, and the constraints, which amount to the requirement that $\underset{\sim}{x}$ be CP, are given by

$$0 \leqslant \left(x_{kk}^{(k-1)}\right)^2 - \left(x_{ij}^{(k-1)}\right)^2 < +\infty \qquad (2 \leqslant k \leqslant n-1, k \leqslant i, j \leqslant n \quad \text{and} \quad (i,j) \neq (k,k)).$$

Because any $x_{pq}^{(r)}$ is a quotient of subdeterminants of $\underset{\sim}{x}$, analytic formulae for the relevant partial derivatives are forbiddingly complicated for arbitrary n, although we did use them for $n = 3$. For larger n we used forward difference quotients to approximate partial derivatives, with a step size equal to the square root of machine precision. Our subroutines were written in terms of variables $x_1, \dots, x_{n^2-1}$, and evaluating the objective and constraint functions and their partial derivatives at one value for $\underset{\sim}{x}$ meant row reducing n^2 different $n \times n$ matrices.

Here is a summary of the examples we obtained from NPSOL, and their implications. We were able to gather evidence relevant to the determination of $c(n)$ and $g(n)$ for $n = 5, 6$, and 7, as well as the verification of the known values $c(3) = g(3) = 9/4$ and $c(4) = g(4) = 4$. Since NPSOL produces approximations to local constrained optima and we are interested in global constrained optima, it was appropriate that we use the package repeatedly with many and various initial points. So we used a random number generator to produce matrices with random entries between -1 and 1, which were then modified so as to satisfy the constraints of the problem at hand. For the length problem we used the Gram-Schmidt procedure, followed by normalization. For the growth problem we performed GECP on a copy of the random matrix to determine the necessary row and column exchanges to make the matrix CP, did those exchanges to the original and then normalized it.

We will call two matrices "essentially the same" if after permuting the rows of one, each pair of corresponding rows are Hadamard equivalent, and the same is true of their transposes. This is weaker than the relation of Hadamard equivalence, defined in (5.13).

For the length problem the behavior of the package was very stable. Every successful run resulted in a matrix which was essentially the same as one of the following:

$$L_3 = \begin{bmatrix} 1 & 1 & 1/2 \\ -1 & 1/2 & 1 \\ 1/2 & -1 & 1 \end{bmatrix} \quad L_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

$$L_5 = \begin{bmatrix} 1 & 1 & 1/3 & 1/2 & -1 \\ -1 & 1 & 1 & 1/3 & 1/2 \\ 1/2 & -1 & 1 & 1 & 1/3 \\ 1/3 & 1/2 & -1 & 1 & 1 \\ 1 & 1/3 & 1/2 & -1 & 1 \end{bmatrix} L_6 = \begin{bmatrix} 1 & -1 & 1 & 1 & 1 & 0 \\ 1 & 1 & -1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & -1 & -1 \\ 1 & 0 & -1 & 1 & 1 & -1 \\ 1 & 1 & 0 & -1 & 1 & -1 \\ 1 & 1 & 1 & 0 & 1 & 1 \end{bmatrix}$$

$$L_7 = \begin{bmatrix} a & -1 & -1 & -b & -1 & 1 & -1 \\ -1 & 1 & -b & 1 & -c & -d & -1 \\ 1 & -1 & c & 1 & a & -c & -c \\ -1 & -1 & d & -1 & -c & -1 & b \\ -b & -a & 1 & 1 & -1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -c & b & -d \\ -1 & -b & 1 & -a & 1 & 1 & -1 \end{bmatrix},$$

where $(a, b, c, d) = ((-2 + \sqrt{7})/3, (3 - \sqrt{7})/2, (-1 + \sqrt{7})/2, (1 + \sqrt{7})/6)$. These values of a, b, c, and d were determined by the requirement that L_7 have the above form and be NO. Since these values occurred on every successful run, we strongly suspect that $c(5) = 121/36 \approx 3.36111$, $c(6) = 5$, and $c(7) = (184 - 35\sqrt{7})/18 \approx 5.07771$. If this value of $c(5)$ is correct, it has the surprising consequence that $c(5) < c(4)$. We have examples which show that $g(5) > 4$ and $g(7) \geqslant 6$, so if the above values for $c(5)$ and $c(7)$ are correct then $c(5) < g(5)$ and $c(7) < g(7)$, hence $c(n) \neq g(n)$ in general.

For the growth problem the behavior of the package was much less stable than for the length problem, and the authors are not certain how to interpret this observed instability. We applied the package with $3 \leqslant n \leqslant 7$ and only with $n = 4$ were the results absolutely uniform: on every successful run we obtained a 4×4 Hadamard matrix. For $n = 3$, most successful runs yielded a matrix essentially the same as L_3 above, but occasionally we obtained a matrix A with $g(A) = 2$. We speculate that this indicates that for $n = 4$ all local extrema for the growth problem are in fact global extrema, but that for $n = 3$ there are several local extrema which are not global.

For $n = 5$, 6 and 7 our successful runs produced matrices A with many different values of $g(A)$. For $n = 6$ the largest value was $g(A) = 5$, and this occurred with A essentially the same as L_6 above. So we are inclined to conjecture that $g(6) = c(6) = 5$.

For $n = 5$, the value $g(A) = 4$ was the smallest produced by a successful run, and it occurred quite often. The largest value we saw occurred with a matrix having growth approximately 4.1325, but the results from all the runs were sufficiently erratic that we decline to conjecture a value for $g(5)$. The example below has growth approximately 4.1296. It was obtained by rounding the entries of one example to 3 digits and adjusting a few to restore the CP property. There are several entries in this example which, if altered slightly, will cause a row exchange to occur during GECP and considerably smaller growth as a result.

$$A_5 = \begin{bmatrix} 1 & .617 & 1 & .778 & -.452 \\ 1 & -.836 & -.452 & -.635 & .999 \\ -.581 & -.997 & .854 & 1 & 1 \\ .451 & -1 & -1 & 1 & -1 \\ -.195 & -1 & .999 & -1 & -.999 \end{bmatrix}.$$

For $n = 7$, successful runs produced values of $g(A)$ ranging from approximately 5.428 to 6. The matrix which gave growth = 6 did not have entries which were recognizably simple. Rounding and adjusting them as we did for A_5, we obtain the following, which has growth approximately 5.996:

$$A_7 = \begin{bmatrix} 1 & -1 & -1 & -1 & 1 & 1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & -1 \\ 1 & -.373 & 1 & -1 & .797 & -.228 & 1 \\ 1 & .196 & -.013 & 1 & 1 & -1 & -1 \\ -1 & .576 & -1 & -1 & 1 & -1 & .026 \\ 1 & 1 & -1 & 1 & 1 & .999 & .999 \\ -1 & 1 & 1 & -1 & .999 & 1 & -.999 \end{bmatrix}.$$

These results suggest several questions, among them the following:

(1) Are our values correct for $c(n)$, $n \leqslant 7$? In particular, is $c(5) < c(4)$?

(2) Is $c(n)$ increasing for $n \geqslant 5$?

(3) Is it true for the length problem that every local extreme point is in fact a global extreme point, for all n?

(4) For what values of n are all local extrema for the growth problem in fact global extrema?

(5) Is $h(A) = c(A) = g(A)$ when A is a normalized orthogonal matrix?

(6) For what values of n does $g(n) = c(n)$?

(7) What is the explanation for the instability we observed in connection with the growth problem?

REFERENCES

J. Brenner and L. Cummings [1972], The Hadamard maximum determinant problem, *American Math Monthly*, 79, 626–630.

A. M. Cohen [1974], A note on pivot size in Gaussian elimination, *Lin. Alg. and Its Applic.*, 8, 361–368.

C. W. Cryer [1968], Pivot size in Gaussian elimination, *Num. Math*, 12, 335–345.

F. R. Gantmacher [1959], Theory of Matrices, Vol. I, Chelsea, New York.

P. E. Gill, W. Murray, and M. Wright [1981], Practical Optimization, Academic Press, New York.

P. E. Gill, W. Murray, M. Saunders, and M. Wright [1983], NPSOL, Systems Optimization Laboratory, Stanford University.

G. H. Golub and C. F. Van Loan [1983], Matrix Computations, Johns Hopkins Univ. Press, Baltimore.

M. J. Hadamard [1893], Rèsolution d'une question relative aux dèterminants, *Bull. des Sciences Math.*, 240–246.

M. Hall [1961], Hadamard matrices of order 16, Jet Propulsion Lab., *Res. Summ.*, 36-10, Vol. 1, 21–26, Pasadena, CA.

______, [1965], Hadamard matrices of order 20, Jet Propulsion Lab., *Tech. Report*, 32–761, Pasadena, CA.

______, [1986], Combinatorial Theory, 2nd ed., John Wiley, New York.

P. Lancaster and M. Tismenetsky [1985], The Theory of Matrices, 2nd ed., Academic Press, New York.

E. S. Lander [1983], Symmetric Designs: An Algebraic Approach, Cambridge Univ. Press, London.

F. J. MacWilliams and N. J. A. Sloane [1977], The Theory of Error-Correcting Codes, North-Holland Pub. Co., New York.

A. W. Marshall and I. Olkin [1979], Inequalities: Theory of Majorization and its Applications, Academic Press, New York.

R. E. A. C. Paley [1933], On orthogonal matrices, *J. Math. Physics*, 12, 311–320.

F. R. Sharpe [1907], The maximum value of a determinant, *Bull. Amer. Math. Soc.*, 14, 121–123.

J. J. Sylvester [1867], Thoughts on inverse orthogonal matrices, *Phil. Mag.*, 34, 461–475.

L. N. Trefethen [1985], Three mysteries of Gaussian elimination, *ACM Signum Newsletter*, October 1985, 2–5.

L. Tornheim [1964], Pivot size in Gauss reduction, Tech. Paper, Chevron Research Co., Richmond, CA.

______, [1965], Maximum third pivot for Gaussian reduction, Tech. Report, Chevron Research Co., Richmond, CA.

______, [1969], A bound for the fifth pivot in Gaussian elimination, Tech. Report, Chevron Research Co., Richmond, CA.

______, [1970], Maximum pivot size in Gaussian elimination with complete pivoting, Tech. Report, Chevron Research Co., Richmond, CA.

J. H. Wilkinson [1961], Error analysis of direct methods of matrix inversion, *Jour. Assoc. Comp. Mach.*, 8, 281–330.

______, [1965], The Algebraic Eigenvalue Problem, Oxford Univ. Press, London.

The Möbius Group and Invariant Spaces of Analytic Functions*

STEPHEN D. FISHER, *Northwestern University*

STEPHEN D. FISHER: I received my bachelor's degree from MIT in 1963. My graduate work was done at the University of Wisconsin, Madison, where I was a student of Frank Forelli. I was an instructor at MIT from 1967 to 1969, at which time I joined the faculty at Northwestern University. My research focuses on the interaction of complex analysis with functional analysis and approximation theory. In addition to a number of research papers, I have written a graduate-level book, *Function Theory on Planar Domains*, John Wiley and Sons, 1983, and an undergraduate text, *Complex Variables*, Wadsworth, 1986.

Introduction. One of the most interesting themes of modern, functional analytic, analysis is the continual recurrence of theorems rooted in classical analysis. By this I mean that what theorems proved some 60, 80, or even 100 or more years ago, say about the growth or range of analytic functions are found to be not only relevant but actually profoundly connected with, and instrumental in, the development of modern Banach-space-oriented investigations. This paper illustrates this principle anew, this time in the investigation of the nature of linear spaces of analytic functions on the unit disc in the complex plane which are invariant, in a sense to be made precise later, under composition with a certain group of linear fractional transformations. The basic results in this area are accessible to anyone who has studied elementary complex variables and who has a smattering of knowledge of what I would term "modern" analysis, that is to say, that portion of analysis that takes as its main interest the study of normed linear spaces, their dualities, and the operators that act on them.

1. The Möbius group and some of its properties. Throughout this paper Δ will denote the open unit disc in the complex plane, $\Delta = \{z\colon |z| < 1\}$. The collection of those analytic functions on Δ that are one-to-one mappings of Δ onto itself form a group, with the group operation of composition. We shall call this group the **Möbius group** and denote it by $\mathbf{G}$. It is elementary that each element of $\mathbf{G}$ is a linear fractional transformation φ of the special form

$$\varphi(z) = \lambda(a - z)/(1 - \bar{a}z), \tag{1}$$

where a is a point of Δ and λ is a constant of absolute value 1; see, for example [**8**, page 198]. Notice, too, that each $\varphi \in \mathbf{G}$ gives a one-to-one orientation-preserving mapping of the unit circle $\mathbf{T} = \{z\colon |z| = 1\}$ onto itself. Define

$$\varphi_a(z) = (a - z)/(1 - \bar{a}z), \qquad a \in \Delta. \tag{2}$$

φ_a interchanges the points a and 0 and is its own inverse: $\varphi_a^{-1} = \varphi_a$.

The elements of $\mathbf{G}$ possess a number of fascinating properties that will play critical roles in our story. The first is the equality

$$|\varphi(z) - \varphi(w)|/\left|1 - \varphi(z)\overline{\varphi(w)}\right| = |z - w|/|1 - z\bar{w}|, \tag{3}$$

*Research and writing supported in part by a grant from the National Science Foundation.

which holds for all points z, w in Δ; this may be established directly by computation. Now divide both sides by $|z - w|$ and let $w \to z$; the result is that

$$(1 - |z|^2)|\varphi'(z)| = 1 - |\varphi(z)|^2. \tag{4}$$

In fact, more than this is true. If f is any analytic function mapping Δ into Δ, then

$$(1 - |z|^2)|f'(z)| \leqslant 1 - |f(z)|^2 \tag{5}$$

and strict inequality holds at all points of the disc Δ *unless* f is an element of **G**. This is the **invariant form of Schwarz's lemma** and its elementary proof follows. Let f be an analytic function mapping Δ into itself, ζ a point of Δ, and $\xi = f(\zeta)$. Define a function g by

$$g(z) = \varphi_\xi(f(z))/\varphi_\zeta(z).$$

g is then an analytic function on Δ since the zero at ζ of φ_ζ in the denominator is cancelled by the zero at ζ of $\varphi_\xi(f(z))$ in the numerator. The maximum principle implies that g also maps Δ into itself. After untangling the relations this yields

$$|f(z) - f(\zeta)||1 - z\bar{\zeta}| \leqslant |1 - f(z)\overline{f(\zeta)}||z - \zeta|. \tag{6}$$

To complete the proof of (5), just divide both sides of (6) by $|z - \zeta|$ and let ζ approach z. The fact that equality holds in (5) only for elements of **G** is again a consequence of the maximum principle. The expression on the left-hand side of (5):

$$Jf(z) = f'(z)(1 - |z|^2) \tag{7}$$

is called the **Möbius invariant derivative of f**, or just the invariant derivative of f for short. To see what this means, let $\varphi \in \mathbf{G}$ and put $w = \varphi(z)$, $h = f \circ \varphi$. Then

$$\begin{aligned} |h'(z)|(1 - |z|^2) &= |f'(\varphi(z))||\varphi'(z)|(1 - |z|^2) \\ &= |f'(\varphi(z))|\left(1 - |\varphi(z)|^2\right) \\ &= |f'(w)|(1 - |w|^2) \end{aligned}$$

as a result of (4). In short,

$$|J(f \circ \varphi)(z)| = |(Jf)(\varphi(z))|. \tag{8}$$

This will have some interesting consequences later, as we shall see.

A second useful property of elements of the Möbius group is their effect on Lebesgue area measure when the change of variables $w = \varphi(z)$ is made. We will use the notation dA for normalized area measure:

$$dA(z) = (1/\pi)\, dx\, dy, \qquad z = x + iy.$$

When we write $\varphi = u + iv$, the usual change of variables formula using the Jacobian gives

$$dA(w) = \left| det \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix} \right| dA(z).$$

However, because φ is analytic the Cauchy-Riemann equations, $u_x = v_y$ and $u_y = -v_x$, imply that the determinant equals $u_x^2 + v_x^2 = |\varphi'|^2$. Consequently,

$$dA(w) = |\varphi'(z)|^2\, dA(z). \tag{9}$$

Let us now apply (4) to (9). We define a measure $d\Sigma$ on Δ by

$$d\Sigma(z) = (1 - |z|^2)^{-2}\, dA(z).$$

(Note that $d\Sigma$ is σ-finite but not finite on Δ.) With respect to each element φ of the group $\mathbf{G}$, the measure $d\Sigma$ then displays the invariance:

$$d\Sigma(w) = d\Sigma(z), \qquad w = \varphi(z). \tag{10}$$

For this reason the measure $d\Sigma$ will be called the **(Möbius) invariant measure on Δ**; the fact that we say **the** Möbius invariant measure is a statement about uniqueness which we will not prove.

On the unit circle $\mathbf{T} = \{z\colon |z| = 1\}$ elements of the Möbius group also produce an important change of variables. If h is a function analytic on a neighborhood of the set $\{z\colon |z| \leqslant 1\}$, φ an element of $\mathbf{G}$ and $\varphi(e^{i\theta}) = e^{it}$, then

$$\begin{aligned} |(h \circ \varphi)'|\, d\theta &= |h'(\varphi(e^{i\theta}))|\,|\varphi'(e^{i\theta})|\, d\theta \\ &= |h'(e^{it})|\, dt. \end{aligned} \tag{11}$$

The Möbius group has strong connections with the non-Euclidean, hyperbolic geometry on the disc Δ. To start, (3) implies that the function

$$d(z, w) = |z - w|/|1 - z\bar{w}|$$

is invariant under $\mathbf{G}$ in the sense that for each $\varphi \in \mathbf{G}$

$$d(\varphi(z), \varphi(w)) = d(z, w).$$

Moreover, (4) may be rewritten in differential notation as

$$|dz|/(1 - |z|^2) = |d\zeta|/(1 - |\zeta|^2), \qquad \zeta = \varphi(z).$$

We define the **hyperbolic length** of a (smooth) curve γ in Δ by

$$L(\gamma) = \int_\gamma |dz|/(1 - |z|^2)$$

and the **hyperbolic area** of a (Lebesgue measurable) set E by

$$A(E) = \iint_E d\Sigma.$$

It follows that each $\varphi \in \mathbf{G}$ is a hyperbolic length and hyperbolic area-preserving transformation. More generally, the inequalities (5) and (6) show that an analytic function f mapping Δ into itself reduces the function d: $d(f(z), f(w)) \leqslant d(z, w)$; reduces hyperbolic length: $L(f(\gamma)) \leqslant L(\gamma)$; and reduces hyperbolic area: $A(f(E)) \leqslant A(E)$. Note, too, that the invariant derivative defined in (7) is nothing but the derivative taken with respect to hyperbolic length.

The hyperbolic geometry in the disc Δ has as its "lines" arcs of circles that meet the unit circle $\mathbf{T}$ twice, each time at a right angle. These are the geodesics of this geometry. The hyperbolic geometry has a metric, the **hyperbolic distance**, defined by,

$$D(z, w) = (1/2)\log[(1 + d(z, w))/(1 - d(z, w))].$$

This is a metric and (3) implies that $D(\varphi(z), \varphi(w)) = D(z, w)$ for all $z, w \in \Delta$ and all elements φ of $\mathbf{G}$. Furthermore, equality holds in the triangle inequality if and only if the three points lie on the same hyperbolic line. Thus, the elements of $\mathbf{G}$ act

as rigid motions relative to this geometry. Moreover, the distance from the origin to the point z, $|z| = r > 0$, is $(1/2)\log[(1 + r)/(1 - r)]$. It is thus infinitely far from the origin, or any other point in Δ, to the unit circle $\mathbf{T}$. More on the hyperbolic geometry of the disc Δ and the group $\mathbf{G}$ can be found in [**25**]; also see [**1**].

2. Spaces of functions invariant under G. The chief concern of this paper is to survey some recent results on linear spaces of analytic functions that are invariant under composition with elements of the Möbius group $\mathbf{G}$.

When I write that $\mathbf{X}$ is **Möbius invariant** I mean by this that $\mathbf{X}$ is a linear space of analytic functions on Δ which is complete in a norm or seminorm, $\|\ \|_{\mathbf{X}}$, and further,

$$\begin{aligned}&\text{for each function } f \text{ in } \mathbf{X} \text{ and each element } \varphi \text{ of } \mathbf{G} \text{ the}\\ &\text{function } f \circ \varphi \text{ also lies in } \mathbf{X} \text{ and satisfies } \|f \circ \varphi\|_{\mathbf{X}} = \|f\|_{\mathbf{X}}.\end{aligned} \tag{12}$$

There are two other restrictions on what precisely constitutes a Möbius invariant space; they will be listed later when they are more relevant. It is more important at this point to list a number of spaces which possess the invariance property (12). The list will be extensive but, of course, not exhaustive.

Example 1. The space H^∞ of bounded analytic functions on Δ with the norm $\|f\| = \sup\{|f(z)|\colon z \in \Delta\}$ satisfies (12). Further, the closed subspace, A, of H^∞ consisting of those functions which are continuous on the closed disc, $\{z\colon |z| \leqslant 1\}$ also satisfies (12). A is called the **disc algebra** and has been the subject of much study in other contexts; see, for instance [**13**].

Example 2. For $1 < p \leqslant \infty$ the space B^p is defined to be all analytic functions g on Δ for which the invariant derivative, Jg, lies in $L^p(\Delta, d\Sigma)$. For $1 < p < \infty$ and after the relation (7) and the definition of $d\Sigma$ are taken into account this amounts to the requirement that the integral

$$\int_\Delta |g'(z)|^p (1 - |z|^2)^{p-2}\, dA(z) \tag{13}$$

be finite. For $p = \infty$ the requirement is that the quantity

$$\sup\left\{|g'(z)|(1 - |z|^2)\colon z \in \Delta\right\} \tag{14}$$

be finite. When $1 < p < \infty$ the space B^p is the **Besov p-space**; they are part of a much more general family of spaces; see [**17**]. For historical reasons, which I will set out below, the space B^∞ is called the **Bloch space**. The seminorm on B^p is the pth root of the integral in (13) if $1 < p < \infty$ and the quantity (14) if $p = \infty$. Note that a constant, and only a constant, has seminorm zero.

The advantage of the initial definition, that the invariant derivative of g belongs to $L^p(\Delta, d\Sigma)$, is that it is immediately clear, using (8) and (10), that

$$\|g\|_p = \|g \circ \varphi\|_p \qquad g \in B^p, \qquad \varphi \in \mathbf{G}, \qquad 1 < p \leqslant \infty.$$

By the way, the reader who is now anticipating the definition of the space B^1 along the lines of the integral (13) will be disappointed. It is an interesting exercise to show that the integral in (13) does not converge if $p = 1$ unless g is constant. The space B^1 will be defined below in Example 4 but it will take some effort to show that it is connected to the spaces B^p as defined here.

Let us spend a second now to show that H^∞ is a subset of B^∞ and, indeed, that $\|g\|_{B^\infty} \leqslant \|g\|_{H^\infty}$ for all $g \in H^\infty$. Let $g \in H^\infty$, set $M = \|g\|_{H^\infty} = \sup\{|g(z)|: z \in \Delta\}$ and define $f = g/M$. Then f maps Δ into Δ and so by (5)

$$(1 - |z|^2)|f'(z)| \leqslant 1 - |f(z)|^2 \leqslant 1$$

for each $z \in \Delta$. Now just multiply through by M to obtain $(1 - |z|^2)|g'(z)| \leqslant M$, which is the desired inequality.

The closed subspace, b^∞, of B^∞ consisting of those functions g for which

$$\operatorname{limit}\{(1 - |z|^2)|g'(z)|: |z| \to 1\} = 0$$

also satisfies (12). Actually, b^∞ is the closure of the polynomials in z in the seminorm (14). b^∞ is called, for notationally clear reasons, the **little Bloch space**.

Example 3. The space B^2 is also known as the **Dirichlet space** and is usually denoted by $\mathbf{D}$. The functions in $\mathbf{D}$ are those for which the area of the image of Δ under f, counting multiplicities, is finite since the integral in (13) becomes (remember, $p = 2$)

$$\int_\Delta |f'|^2\, dA = \int_{f(\Delta)} 1\, dA.$$

Note also that the Dirichlet seminorm is also given by the expression

$$\|f\| = \left\{\sum_{k=1}^{\infty} k|a_k|^2\right\}^{1/2}, \qquad f(z) = \sum_{k=0}^{\infty} \partial_k z^k$$

The space $\mathbf{D}$ is a semi-Hilbert space meaning that it has an inner product

$$(f, g)_{\mathbf{D}} = \int_\Delta f'\overline{g'}\, dA \tag{15}$$

which satisfies all the usual requirements of an inner product except that it is possible that $(f, f)_{\mathbf{D}}$ can be zero without f being zero; this happens, as we've seen, if and only if f is a constant. Later we shall discuss the possibility of the existence of other, intrinsically different Möbius invariant semi-Hilbert spaces.

Example 4. The space B^1 is defined to be all those analytic functions f on Δ which have a representation as

$$f(z) = \sum_{k=1}^{\infty} c_k\varphi_k, \quad \text{where} \quad \varphi_k \in \mathbf{G} \quad \text{and} \quad \sum_{k=1}^{\infty} |c_k| < \infty. \tag{16}$$

Since a function f could conceivably have several such representations, we define the norm on B^1 by the rule

$$\|f\| = \inf\left\{\sum_{k=1}^{\infty} |c_k| : (16) \text{ holds}\right\}. \tag{17}$$

The invariance property (12) is proved by the following argument. If $f \in B^1$ and $\varphi \in \mathbf{G}$, then $f \circ \varphi = \Sigma c_k\varphi_k \circ \varphi$. Since $\mathbf{G}$ is a group, $\varphi_k \circ \varphi$ again lies in $\mathbf{G}$ so we obtain $\|f \circ \varphi\| \leqslant \Sigma|c_k|$. By taking the infimum over all representations of f we find that $\|f \circ \varphi\| \leqslant \|f\|$. However, because $\mathbf{G}$ is a group, the same inequality is true with φ^{-1} in place of φ; equivalently, $\|f\| \leqslant \|f \circ \varphi\|$. Thus, equality holds. The functions

$\varphi \in \mathbf{G}$ form the "atoms" out of which the space B^1 is built. The interested reader could spend a minute and prove that $\|\varphi\| = 1$ for each $\varphi \in \mathbf{G}$. (It's clear that φ has norm at most 1; the issue is to show that it cannot be less than 1. One approach is to show that each $f \in B^1$ lies in the disc algebra A and its norm in A is no more than its norm in B^1.)

Remark. The familiar Hardy H^p spaces are defined for $0 < p < \infty$ by the requirement that

$$\|f\|_{H^p} = \sup\left\{\left[(1/2\pi)\int_0^{2\pi}|f(re^{i\theta})|^p\,d\theta\right]^{1/p} : 0 < r < 1\right\}$$

be finite. It is worth pointing out explicitly that *none* of these H^p spaces satisfy (12). Indeed, it is not a difficult computation (see [7]) to show that

$$\sup\{\|f \circ \varphi_a\|_{H^p} \colon f \in H^p, \|f\| \leqslant 1\} \geqslant (1 - |a|^2)^{-1/p}$$

and hence (12), or anything like it, does *not* hold for H^p.

Example 5. Those analytic functions h on Δ for which h' lies in the Hardy space H^1 also satisfy (12). The seminorm on this space is

$$\|h\| = (1/2\pi)\int_0^{2\pi}|h'(e^{i\theta})|\,d\theta$$

and the invariance is a result of the change of variables formula (11). The functions in this space are exactly those that map the unit circle $\mathbf{T}$ onto a rectifiable curve of finite length; see [7].

3. Basic properties of Möbius invariant spaces. Let us begin our look at Möbius invariant spaces by identifying the largest and the smallest ones. We first list an additional, most natural, property that each Möbius invariant space is required to possess.

There is some continuous linear functional ℓ on $\mathbf{X}$ which is also continuous in the topology of uniform convergence on compact subsets of Δ. (18)

That is, if $\{f_n\}$ is a sequence of elements of $\mathbf{X}$ which converges uniformly on compact subsets of Δ to 0, then $\ell(f_n) \to 0$. For instance, $\ell(f) = f'(0)$ satisfies (18) on each of the spaces listed in Examples 1–5. With the modest requirements, (12) and (18), we are now ready to prove our first theorem, a fundamental result of L. A. Rubel and R. M. Timoney [24].

THEOREM 1. *Let* $\mathbf{X}$ *satisfy* (12) *and* (18). *Then* $\mathbf{X}$ *is a linear subspace of the Bloch space* B^∞ *and the embedding is continuous: there is a constant* C *depending on* $\mathbf{X}$ *but not* h *such that*

$$\|h\|_{B^\infty} \leqslant C\|h\|_{\mathbf{X}} \quad \text{for all } h \in \mathbf{X}. \tag{19}$$

Proof. The linear function ℓ has the form $\ell(h) = \int_0^{2\pi} h(re^{i\theta})G(re^{i\theta})\,d\theta$, where $G(z)$ is analytic on a neighborhood of the set $|z| \geqslant r$, including at ∞, for some r, $0 < r < 1$; see [23]. Let $G(z) = \sum_{n=0}^{\infty} G_n z^{-n}$ be the power series expansion of G about ∞ and let N be any integer such that $G_N \neq 0$. Replace h by $e^{-iNt}h(ze^{it})$ and

then integrate with respect to t from 0 to 2π. Using (12) we obtain the estimate $r^{2N}|h_N G_N| \leqslant C\|h\|_{\mathbf{X}}$ where C is the norm of ℓ on $\mathbf{X}$. This is nearly what we want. If $N = 1$, then this inequality shows that $|h'(0)| \leqslant C'\|h\|_{\mathbf{X}}$ for some constant C' independent of $h \in \mathbf{X}$. Replace h by $h \circ \varphi_a$; we are led to the inequality $(1 - |a|^2)|h'(a)| \leqslant C'\|h\|_{\mathbf{X}}$, valid for each $a \in \Delta$. This, of course, implies that $\mathbf{X}$ is a subset B^∞ and we are done. If $N > 1$, then similar, but more elaborate, computations show that $\mathbf{X}$ is continuously embedded in B^∞. (The reader can easily work out the omitted case $N = 0$ and the case $N = 2$.) □

This is a good point to explain why the functions in B^∞ are called Bloch functions. The starting point is the 1924 theorem of A. Bloch [**5**], which states that an analytic function f in Δ, normalized by the condition $f'(0) = 1$, must cover, in a one-one fashion, a disc in its range of at least a certain minimum size. This is one of a number of beautiful and profound results about the range of an analytic function, including the theorems of Koebe and Picard, which, by the way, is a consequence of Bloch's theorem. In the course of giving another proof of Bloch's result and deriving a numerical lower bound for the minimum size of this disc, E. Landau [**14**] showed by a series of elementary manipulations that attention could be restricted to functions f and satisfied the growth condition $|f'(z)|(1 - |z|^2) \leqslant 1$. Of course, mathematicians being what they are, this last condition was ultimately recognized as defining the unit ball of some space and "Bloch functions" as entities unto themselves were born. More on Bloch functions and their fascinating properties can be found in [**2**] and a proof of Bloch's theorem in [**6**]. We now return to our primary interest, Möbius invariant spaces.

The third and final restriction that we impose in order that a space be Möbius invariant requires us to topologize the Möbius group $\mathbf{G}$. We do this by requiring that the mapping $(\lambda, a) \to \lambda\varphi_a$ of $\mathbf{T}\mathrm{x}\Delta$ onto $\mathbf{G}$ is a homeomorphism. This means, basically, that the disc and the unit circle each have their usual topology but you can't reach one from the other. In order that a space $\mathbf{X}$ be Möbius invariant we require that (12) and (18) hold and, in addition, the final restriction:

for each $f \in \mathbf{X}$ the mapping $\varphi \to f \circ \varphi$ of the group $\mathbf{G}$ into $\mathbf{X}$ is norm continuous. (20)

By the way, for the spaces H^∞ and B^∞, (20) is not correct. Luckily, all is not lost. Both of these spaces are dual spaces and the mapping in (20) is continuous from $\mathbf{G}$ into them when they are given the weak-* topology. So we can modify (20) for dual spaces to require only that the mapping in (20) be continuous from $\mathbf{G}$ into $\mathbf{X}$ with the weak-* topology.

We are now ready to identify B^1 as the smallest Möbius invariant space.

THEOREM 2. *If* $\mathbf{X}$ *is a nontrivial Möbius invariant space, then* $\mathbf{X}$ *contains* B^1. *Indeed, the containment is continuous in the sense that*

$$\|f\|_{\mathbf{X}} \leqslant C\|f\|_{B^1} \quad \textit{for all } f \in B^1 \tag{21}$$

where C *is a constant depending on* $\mathbf{X}$ *but not on* f.

Proof. We shall first show that $\mathbf{X}$ contains each of the functions z^n, $n = 1, 2, \ldots$. Since $\mathbf{X}$ is nontrivial it contains a function F which is not constant. Let $F(z) = \sum_{k=0}^{\infty} b_k z^k$ be the power series expansion of F about the origin and choose a positive

integer N such that $b_N \neq 0$. The function g defined by

$$g(z) = (1/2\pi)\int_0^{2\pi} F(ze^{i\theta})e^{-iN\theta}\,d\theta, \qquad z \in \Delta$$

lies in $\mathbf{X}$ since the integrand is a continuous, $\mathbf{X}$-valued function of θ by (20). On the other hand, the value of the integral is most clearly $b_N z^N$. This shows that z^N lies in $\mathbf{X}$. It follows from (12) that $(\varphi_a)^N$ lies in $\mathbf{X}$ for each $a \in \Delta$ and repeating the above argument with this function in place of F we see that z^n lies in $\mathbf{X}$. (This may involve choosing an appropriate $a \in \Delta$.) The rest is now easy. Since z lies in $\mathbf{X}$ so does each $\varphi \in \mathbf{G}$, again by (12) and

$$\|\varphi\|_{\mathbf{X}} = \|z \circ \varphi\|_{\mathbf{X}} = \|z\|_{\mathbf{X}} = C.$$

Now let $f = \Sigma c_k \varphi_k$ be any element of B^1. We then compute

$$\begin{aligned}\left\|\sum c_k\varphi_k\right\|_{\mathbf{X}} &\leqslant \sum |c_k| \|\varphi_k\|_{\mathbf{X}} \\ &= C\sum |c_k|.\end{aligned}$$

Taking the infimum over all representations of f we obtain (21). □

COROLLARY 3. *A Möbius invariant space contains the polynomials.*

Indeed, it is not hard to show that the polynomials are dense in a Möbius invariant space. (Or weak-* dense in the case of a Möbius invariant dual space.)

There is a problem here that we cannot ignore—the space B^1 itself is not obviously Möbius invariant. We've certainly shown that it satisfies (12). (18) is also easy since if $f = \Sigma c_k \varphi_k$, then f is the absolutely convergent sum of functions of unit norm in the disc algebra and so lies in the disc algebra. In particular, for any point $z_0 \in \Delta$, $|f(z_0)| \leqslant \Sigma |c_k|$; this gives (18). The problem is, of course, with (20). While it is true that B^1 *does* satisfy (20) the proof would take us too far afield so we shall just content ourselves with stating this as a proposition; see [4] for the details.

PROPOSITION 4. *The space B^1 satisfies* (20) *and so is the minimal Möbius invariant space.*

Let's spend a paragraph or two here to show how the space B^1 fits into the family of spaces B^p as defined in Example 2. First, it is not too difficult to establish that for $1 < p < \infty$ a function f lies in B^p if and only if

$$\int_\Delta |f''(z)|^p (1 - |z|^2)^{2p-2}\,dA(z) \tag{22}$$

is finite. The trade-off of one derivative of f for an extra power $1 - |z|^2$ is basically just a consequence of Cauchy's formula. The disadvantage of formulating membership in B^p this way is that it is not at all obvious from (22) that B^p is Möbius invariant. Nonetheless, (22) does permit us to pass to the case $p = 1$ and connect the space B^1 defined in Example 4 with the B^p spaces.

THEOREM 5. *The space B^1 is identical with the set of those analytic functions f on Δ for which f'' lies in $L^1(\Delta, dA)$.*

Proof. One inclusion is elementary. A computation establishes the inequality

$$\int_\Delta |\varphi_a''(z)|\,dA(z) \leqslant 4, \qquad a \in \Delta.$$

From this it is immediate that each element of B^1 lies in $L^1(\Delta, dA)$ and the norm of its second derivative in L^1 is no more than 4 times its norm in B^1. On the other hand, let h be a polynomial with a zero of order two or more at the origin. Another elementary computation establishes the formula

$$h(z) = \int_\Delta h''(a)\varphi_a(z)/\bar{a}\, dA(a).$$

The integrand is a continuous B^1-valued function of a. The measure $(1/\bar{a})\, dA$ simplifies to $e^{it}\, dt\, dr$, $a = re^{it}$, and this can be used in an elementary fashion to obtain the estimate

$$\|h\|_{B^1} \leqslant C\|h''\|_{L^1(\Delta, dA)},$$

where C is a constant independent of h. The polynomials in z are dense in the analytic functions that lie in $L^1(\Delta, dA)$ and they are also dense in B^1, since B^1 is Möbius invariant. This establishes the other inclusion and finishes the proof. □

4. Dualities, representations, uniqueness. We begin this section with the general thought of finding the dual space of a Möbius invariant space. There are two "principles" that should be kept in mind when searching for dualities among Möbius invariant spaces:

—a Möbius invariant space should have a Möbius invariant dual

—the pairing used for the semi-inner product in the Dirichlet space **D**

$$\langle\langle f, g\rangle\rangle = \int f'\overline{g'}\, dA \tag{23}$$

is Möbius invariant and hence offers the possibility of a Möbius invariant pairing for other pairs of spaces. One point is worth noting here about the integral in (23)—generally, it is not absolutely convergent. We usually need to take the pairing in the limiting sense:

$$\langle\langle f, g\rangle\rangle = \text{limit}\left\{(1/2\pi)\int_0^{1-\varepsilon}\int_0^{2\pi} f'(re^{it})\overline{g'(re^{it})}\, dt\, r\, dr\colon\ \varepsilon \downarrow 0\right\}.$$

The fact that (23) is invariant under **G** is easy to prove, assuming say, that the functions f and g are analytic on a neighborhood of $\{z\colon\ |z| \leqslant 1\}$. For if $\varphi \in \mathbf{G}$, then

$$\begin{aligned}\langle\langle f\circ\varphi, g\circ\varphi\rangle\rangle &= \int_\Delta (f\circ\varphi)'\overline{(g\circ\varphi)'}\, dA\\ &= \int_\Delta f'(\varphi(z))\overline{g'(\varphi(z))}|\varphi'(z)|^2\, dA(z)\\ &= \int_\Delta f'(w)\overline{g'(w)}\, dA(w) = \langle\langle f, g\rangle\rangle\end{aligned}$$

by invoking (9). By the way, the pairing (23) can be realized another way, too, on the unit circle **T** by

$$\langle\langle f, g\rangle\rangle = (1/2\pi)\int_0^{2\pi} e^{-i\theta} f(e^{i\theta})\overline{g'(e^{i\theta})}\, d\theta.$$

Once again the integral is not absolutely convergent, in general, and we have to take

the integral in some sort of limiting sense, say

$$\langle\langle f, g\rangle\rangle = \operatorname{limit}\left\{(1/2\pi)\int_0^{2\pi} e^{-i\theta} f(re^{i\theta})\overline{g'(re^{i\theta})}\,d\theta\colon r \uparrow 1\right\}.$$

One other note about (23): if either of f or g is a constant, then the value of $\langle\langle\, ,\rangle\rangle$ is zero. Therefore, any dualities established with $\langle\langle\, ,\rangle\rangle$ must be taken modulo constants. Let's put the pairing $\langle\langle\, ,\rangle\rangle$ to work with $f \in B^\infty$ and $g = \varphi \in \mathbf{G}$. Another computation produces

$$\langle\langle f, \varphi_a\rangle\rangle = -(1 - |a|^2)f'(a), \qquad a \in \Delta. \tag{24}$$

(Just check this for $f(z) = z^m$, which is very direct, to be convinced.) The right-hand side, of course, is just the quantity that is used in computing the B^∞ norm. Meanwhile, the functions $\{\varphi_a\colon a \in \Delta\}$ are the "atoms" which make up the space B^1. Hence, it should be no surprise that B^∞ and B^1 are in some sort of duality. The results are set forth here; proofs are contained in [**4**].

THEOREM 6. *The dual space of B^∞ is B^1 and the dual space of B^1 is B^∞. In both cases the pairing is effected by* (23) *and the correspondences are isometric isomorphisms.*

Actually, the successive dualities involving, in turn, a "little oh" space, an L^1 space, and finally, a "big oh" space are a well known phenomena in functional analysis, the classic case being the successive dualities of c_0, ℓ^1, and ℓ^∞. However, this type of relationship has already been identified as occurring in spaces of analytic functions on Δ; see, for instance [**22**] and [**27**].

Other dualities among Möbius invariant spaces are also effected by (23). For example, the dual space of B^p is B^q, $p^{-1} + q^{-1} = 1$, and $1 < p < \infty$. However, the correspondence is not isometric in this case. Using the pairing (23) it is also possible to construct, for example, a Möbius invariant predual of H^∞ (the conventional one is not Möbius invariant), and Möbius invariant duals of other Möbius invariant spaces.

While on the subject of duality let us turn here to the topic mentioned in Example 3, that is, the possible existence of a Möbius invariant semi-Hilbert space other than the Dirichlet space **D**. The answer is contained in the next theorem and the remarks that follow it.

THEOREM 7. *Let* $\mathbf{X}$ *be a Möbius invariant semi-Hilbert space. Then the seminorm of* $\mathbf{X}$ *is a constant multiple of the Dirichlet seminorm.*

Proof. Let (f, g) be the semi-inner product in $\mathbf{X}$. The Möbius invariance of the seminorm of $\mathbf{X}$ implies that $(f, g) = (f \circ \varphi, g \circ \varphi)$ for all f, $g \in \mathbf{X}$ and all $\varphi \in \mathbf{G}$. Take $\varphi(z) = e^{it}z$, $f(z) = z^k$, and $g(z) = z^m$, $m \neq k$, and integrate with respect to t from 0 to 2π. This implies that

$$(z^k, z^m) = 0, \qquad m \neq k. \tag{25}$$

Next, take $f(z) = g(z) = z$ and $\varphi = \varphi_r$, $-1 < r < 1$. Making use of (25), we obtain

$$\lambda = (z, z) = (\varphi_r, \varphi_r) = r^2(1,1) + (1 - r^2)^2 \sum r^{2k-2}(z^k, z^k).$$

Expand the right-hand side in powers of r^2. Each coefficient, except the 0th is zero.

This produces the series of equations

$$(1,1) = 0 \quad \text{and} \quad (z^k, z^k) - 2(z^{k-1}, z^{k-1}) + (z^{k-2}, z^{k-2}) = 0, \qquad k \geqslant 2.$$

These equations yield $(z^k, z^k) = k\lambda$, $k = 1, 2, \ldots$. That is, the seminorm in $\mathbf{X}$ is given by

$$\{\|f\|_{\mathbf{X}}\}^2 = (f, f) = \lambda \sum_{k=1}^{\infty} k|a_k|^2, \qquad f(z) = \sum_{k=0}^{\infty} a_k z^k.$$

Hence, the seminorm of $\mathbf{X}$ is λ times the seminorm of $\mathbf{D}$ and the theorem is proved.

Actually, a substantially stronger result than Theorem 7 is true. If $\mathbf{X}$ is a semi-Hilbert space for which (18) and (20) hold and if there is a constant C such that $(f \circ \varphi, f \circ \varphi) \leqslant C(f, f)$ for all $\varphi \in \mathbf{G}$, then the seminorm in $\mathbf{X}$ is equivalent to that of $\mathbf{D}$ and so $\mathbf{X}$ is $\mathbf{D}$; see [3]. This last result tells us that there is only one Möbius invariant semi-Hilbert space. Whether similar uniqueness theorems hold for other types of Möbius invariant spaces is an intriguing question.

Representations of $\mathbf{G}$. Let φ be an element of $\mathbf{G}$ and define a mapping C_φ on analytic functions by $C_\varphi(f) = f \circ \varphi$. The connection between the group $\mathbf{G}$ and the family of mappings C_φ given by

$$\tau(\varphi) = C_\varphi \tag{26}$$

is an anti-isomorphism of groups, the image being a subgroup of the group of continuous linear mappings of the Möbius invariant space $\mathbf{X}$ into itself. As a matter of fact, each C_φ is an isometry of $\mathbf{X}$ and, in particular, of the semi-Hilbert space $\mathbf{D}$ onto itself. Hence, τ is a **representation** of $\mathbf{G}$ as a group of isometries of a (semi)-Hilbert space. There are, however, many other representations of $\mathbf{G}$ as a group of isometries of some Hilbert space. Another such is as unitary mappings of the Hardy space H^2, given by

$$V_\varphi(h) = (h \circ \varphi)(\varphi')^{1/2}. \tag{27}$$

The subject of the representation of a group as a group of isometries of some Hilbert space is an important part of harmonic analysis; see, for example [**10**] and [**15**].

5. Hankel operators. A series of recent results draw further connections between Möbius invariant spaces and other parts of analysis, more specifically portions of operator theory. Recall that a **Hankel matrix** H is a matrix, finite or infinite, whose entries are constant on the diagonals from upper right to lower left; that is, H_{ij} is a function only of $i + j$. Hankel matrices arise in many contexts in analysis; see, for instance [**18**] and [**20**]. One of the most beautiful connections between complex analysis and Hankel operators is the following.

KRONECKER'S THEOREM. *Let* $\{H_k\}_0^\infty$ *be a sequence of complex numbers. The Hankel matrix* $\mathbf{H} = (H_{i+j})$ *has finite rank if and only if the analytic function* $F(z) = \sum_{k=0}^{\infty} H_k/z^k$ *is rational. If this is so, then the rank of the matrix* $\mathbf{H}$ *is equal to the degree of* F.

A proof of Kronecker's theorem and more on Hankel matrices may be found in [**9**]. A **Hankel operator** is an operator on a Hilbert space whose matrix is a Hankel

matrix. One particularly well understood setting for this theory is on the Hardy space H^2, considered as a closed subspace of $L^2(\mathbf{T}, d\theta)$. Let P be the orthogonal projection of $L^2(\mathbf{T}, d\theta)$ onto H^2. P is given by

$$P\left(\sum_{-\infty}^{\infty} a_k e^{ik\theta}\right) = \sum_{k=0}^{\infty} a_k e^{ik\theta}. \tag{28}$$

Suppose that f is an analytic function on Δ. The **Hankel operator induced by f** on H^2 is defined by

$$H_f(g) = (I - P)\bar{f}g. \tag{29}$$

Explicitly, H_f is given by the formula

$$H_f(g)(z) = (1/2\pi i)\int_{\mathbf{T}}\{\overline{f(z)} - \overline{f(\xi)}\}g(\xi)\{\xi - z\}^{-1}\,d\xi. \tag{30}$$

To see why H_f is a Hankel operator we compute its matrix. The functions $e_k = e^{ikt}$, $k = 0, 1, 2, \ldots$ are an orthonormal basis for H^2 and the functions e_{-k}, $k = 1, 2, 3, \ldots$ an orthonormal basis for the range of H_f. Let $f = \sum_{j=0}^{\infty} b_j e^{ijt}$ be the Fourier expansion of f. The (k, n)th entry in the matrix of H_f is given by

$$\left(H_f(e_k), e_{-n}\right)_{L^2} = b_{n+k}, \quad n \geqslant 1, k \geqslant 0. \tag{31}$$

This shows that H_f is indeed Hankel.

What is the connection between these operators and Möbius invariant spaces? The answer to this question lies in finding the relationship between the operators H_f and $H_{f\circ\varphi}$, where φ is an element of $\mathbf{G}$. We begin by noting that the operator V_φ defined in (27) commutes with P and, moreover, is connected with the operator M_u of multiplication by the function u by the rule $M_{u\circ\varphi} = V_\varphi M_u V_{\varphi^{-1}}$; the latter formula is almost a tautology. Let $\psi = \varphi^{-1}$. We then compute

$$\begin{aligned} H_{f\circ\varphi} &= (I - P)M_{\bar{f}\circ\varphi} = \left(V_\varphi V_\psi - P\right)V_\varphi M_{\bar{f}} V_\psi = \left(V_\varphi M_{\bar{f}} - PV_\varphi M_{\bar{f}}\right)V_\psi \\ &= \left(V_\varphi M_{\bar{f}} - V_\varphi P M_{\bar{f}}\right)V_\psi = V_\varphi(I - P)M_{\bar{f}}V_\psi. \end{aligned}$$

That is,

$$H_{f\circ\varphi} = V_\varphi H_f V_\varphi^{-1}. \tag{32}$$

Therefore, the operator $H_{f\circ\varphi}$ is **unitarily equivalent** to H_f. In particular, any property of H_f that is preserved by unitary equivalence of operators (e.g., boundedness, compactness, Hilbert-Schmidt, etc.) will also be possessed by $H_{f\circ\varphi}$. Put into different words, if $\mathbf{X}$ is the space of those analytic functions f on Δ for which the operator H_f possesses property **P**, **P** a unitary invariant of operators, then $\mathbf{X}$ is (at least a good candidate to be) a Möbius invariant space. We show below how this observation can be used together with Theorem 7 to provide a short, direct proof of the characterization of Hilbert-Schmidt Hankel operators.

Let us continue our brief look at Hankel operators by characterizing those that are bounded (= continuous) and those that are compact. These conditions involve the space BMOA. This space consists of those functions f in the Hardy space H^2 which satisfy the additional condition that the quantity

$$\sup\left\{(1/2\pi)\int_0^{2\pi}|f(e^{it})|^2 P_a(e^{it})\,dt - |f(a)|^2 : a \in \Delta\right\} \tag{33}$$

is finite, where P_a is the Poisson kernel for $a \in \Delta$:

$$P_a(e^{it}) = (1 - |a|^2)/|1 - e^{it}\bar{a}|^2.$$

The space VMOA is defined as the subspace of BMOA consisting of those functions for which the quantity in the curly brackets in (33) goes to zero as $|a| \to 1$. The reader should take note of the fact that both BMOA and VMOA are Möbius invariant spaces, with the square root of the quantity in (33) as the norm. The "classical" theorems which characterize boundedness and compactness of the Hankel operator H_f, then, are due to Z. Nehari [**16**] and P. Hartman [**12**], respectively, and are simply stated.

THEOREM 8. *H_f is bounded on H^2 if and only if $f \in BMOA$. H_f is compact on H^2 if and only if $f \in VMOA$.*

Now we turn to Hilbert-Schmidt operators and the connection with Möbius invariant spaces. A linear operator T mapping a Hilbert space K into itself is **Hilbert-Schmidt** if there is an orthonormal basis $\{u_j\}_1^\infty$ of K such that $\sum_{j=1}^\infty \|Tu_j\|^2$ is finite. Further, if T is Hilbert-Schmidt, then this sum is finite for any orthonormal basis and the value of the sum is independent of the choice of the orthonormal basis. The quantity

$$[T, S] = \sum_{j=1}^{\infty} (Tu_j, Su_j)_{\mathbf{K}}$$

defines an inner product on the linear space of Hilbert-Schmidt operators, turning it into a Hilbert space; see [**11**]. The observations above, in particular (32), then set the stage for the following theorem.

THEOREM 9. *The Hankel operator H_f is Hilbert-Schmidt if and only if the function f is in the Dirichlet space. Further, the Hilbert-Schmidt norm of H_f is equal to the Dirichlet norm of f.*

Proof. We define $\mathbf{X}$ to be all those analytic functions f on Δ for which the operator H_f is Hilbert-Schmidt. A seminorm is defined on $\mathbf{X}$ by $\|f\| = \|H_f\|$, where the latter norm is the Hilbert-Schmidt norm of the operator H_f. It is then apparent from (32) that $\mathbf{X}$ is a semi-Hilbert space which satisfies the Möbius invariant property (12). Further, it is not difficult to show that properties (18) and (20) also hold for $\mathbf{X}$. Theorem 7 then shows that $\mathbf{X}$ coincides with the Dirichlet space provided it contains a non-constant function. However, the Hankel operator induced by $F(z) = z$ is simply $H_F(g)(z) = g(0)$. Acting on the orthonormal basis $\{z^k\}_0^\infty$ of H^2 it produces in turn the constant functions $1, 0, 0, \ldots$ and so clearly it is Hilbert-Schmidt with Hilbert-Schmidt norm 1. (The reader will note that, in fact, (32) then shows that H_φ has rank one for all $\varphi \in \mathbf{G}$.) □

More recently, V. V. Peller [**18**], R. Rochberg [**21**], S. Semmes [**26**], and others have made further investigations of the properties of the operator H_f, including theorems about containment of the operator H_f in certain ideals of compact operators in the space of operators. One interesting aspect of these results is that the function spaces so arrived at are the B^p spaces defined in Example 2.

I close by extending my thanks to Professors Jonathan Arazy of Haifa University, Israel and Jaak Peetre of the University of Lund, Sweden. They introduced me to the subject of Möbius invariant spaces of analytic functions and our research collaboration has been fruitful and enjoyable. This paper is due in no small part to the interest and knowledge of this subject which I gained from them.

REFERENCES

1. L. V. Ahlfors, Conformal Invariants, McGraw-Hill, New York, N.Y. 1973.
2. J. M. Anderson, J. Clunie, and Ch. Pommerenke, On Bloch functions and normal functions, *J. reine angew. Math.*, 270 (1974) 12–37.
3. J. Arazy and S. Fisher, The uniqueness of the Dirichlet space among Möbius invariant Hilbert spaces, *Ill. J. Math.*, 29 (1985) 449–462.
4. J. Arazy, S. Fisher, and J. Peetre, Möbius invariant function spaces, *J. reine angew. Math.*, 363 (1985) 110–145.
5. A. Bloch, Les theorems de M. Valiron sur les fonctions entieres et la theorie de l'uniformisation, *Comptes Rendus*, Paris, 178 (1924) 2051–2052, and *Annales de la Faculte des Sciences de l'Universite de Toulouse*, Ser. 3, 17 (1925) 1–22.
6. J. A. Cima, The basic properties of Bloch functions, *Internat. J. Math. and Math. Sci.*, 2 (1979), 369–413.
7. P. Duren, The Theory of H^p Spaces, Academic Press, New York, 1970.
8. S. Fisher, Complex Variables, Wadsworth, Belmont, CA., 1986.
9. F. R. Gantmacher, The Theory of Matrices, vol II, Chelsea, New York, 1959.
10. I. M. Gel'fand, M. I. Graev, N. Ya. Vilenkin, Generalized Functions, vol. 5, Academic Press, New York, 1966.
11. I. C. Gohberg and M. G. Kreín, Introduction to the Theory of Linear Nonselfadjoint Operators, Translations of Math. Monographs, vol. 18, Amer. Math. Soc., Providence, R.I., 1969.
12. P. Hartman, Completely continuous Hankel matrices, *Proc. Amer. Math. Soc.*, 9 (1958) 862–866.
13. K. Hoffman, Banach Spaces of Analytic Functions, Prentice-Hall, Englewood Cliffs, N.J. 1962.
14. E. Landau, Uber die Blochsche konstante und zwei verwandte Weltkonstanten, *Math. Zeit.*, 30 (1929) 608–634.
15. S. Lang, $SL_2(R)$, Addison-Wesley, Reading, MA, 1975.
16. Z. Nehari, On bounded bilinear forms, *Annals of Math.*, 65 (1957) 153–162.
17. J. Peetre, New Thoughts on Besov Spaces, Duke University Math. Series 1, Duke University Press, Durham, N.C., 1976.
18. V. V. Peller, Hankel operators of class S_p and applications (rational approximation, Gaussian processes, the majorant problem for operators), *Math. USSR Sbornik*, 41 (1982) 443–479.
19. S. C. Power, Hankel operators on Hilbert space, *Bull. London Math. Soc.*, 12 (1980) 422–442.
20. S. C. Power, Hankel Operators on Hilbert Space, Research Notes in Mathematics 64, Pittman, Boston, 1982.
21. R. Rochberg, Trace ideal criteria for Hankel operators and commutators, *Indiana Univ. Math. J.*, 31 (1982) 913–925.
22. L. A. Rubel and A. L. Shields, The second duals of certain spaces of analytic functions, *J. Australian Math. Soc.*, 11 (1970) 276–280.
23. L. A. Rubel and B. A. Taylor, Functional analysis proofs of some theorems in function theory, *Amer. Math. Monthly*, 76 (1969) 483–489.
24. L. A. Rubel and R. M. Timoney, An extremal property of the Bloch space, *Proc. Amer. Math. Soc.*, 75 (1979) 45–50.
25. H. Schwerdtfeger, Geometry of Complex Numbers, Dover, New York, N.Y., 1979.
26. S. Semmes, Trace ideal criterion for Hankel operators, $0 < p < 1$, *Integral Equations and Operator Theory*, 7 (1984) 241–281.
27. A. L. Shields and D. L. Williams, Bounded projections, duality, and multipliers in spaces of bounded analytic functions, *Trans. Amer. Math. Soc.*, 162 (1971) 287–302.

On the Zeros of the Taylor Polynomials Associated with the Exponential Function

BRIAN CONREY and AMIT GHOSH, *Oklahoma State University, Stillwater*

BRIAN CONREY did his undergraduate work at Santa Clara University and his graduate work at the University of Michigan (with a year at Cambridge University). He did postdoctoral work at the University of Illinois and the Institute for Advanced Study, Princeton. He is currently a Sloan Fellow.

AMIT GHOSH did his undergraduate work at Imperial College, London and his Ph.D. at the University of Nottingham (with a year at the University of Illinois, Urbana) and postdoctoral work at the Institute for Advanced Study, Princeton. His main interest is in the analytic theory of L-functions.

While investigating a certain mean-value associated with the zeros of the nth derivative of the Riemann zeta-function [**2**] we obtained for each nonnegative integer n a formula with a constant factor

$$\alpha_n := n + 1 - \sum_{\nu=1}^{n} e^{-z_\nu},$$

where the complex numbers z_ν are the roots of the polynomial

$$E_n(z) := 1 + z + \frac{z^2}{2!} + \cdots + \frac{z^n}{n!}.$$

Thus $\alpha_0 = 1$, $\alpha_1 = 2 - e$, $\alpha_3 = 3 - 2e\cos 1$, and if

$$r = (\sqrt{2} + 1)^{1/3} \quad \text{and} \quad s = (\sqrt{2} - 1)^{1/3},$$

then

$$\alpha_3 = 4 - e^{2+r-s} - 2e^{(1+(s-r)/2)}\cos\frac{\sqrt{3}}{2}(r + s).$$

The numbers z_ν have been well studied, and for a given n most of them have real part smaller than $-kn$ for some positive constant k. Thus it is natural to expect that α_n should grow exponentially with n. We computed the first few α_n and

Supported in part by an NSF grant.

quickly changed our expectations:

$$
\begin{array}{llll}
\alpha_0 = 1 & & \alpha_8 = & -0.0000062064\ldots \\
\alpha_1 = & -0.7182818284\ldots & \alpha_9 = & -0.0000018672\ldots \\
\alpha_2 = & +0.0626121201\ldots & \alpha_{10} = & -0.0000004703\ldots \\
\alpha_3 = & +0.0120619221\ldots & \alpha_{11} = & -0.0000000989\ldots \\
\alpha_4 = & +0.0019468374\ldots & \alpha_{12} = & -0.0000000153\ldots \\
\alpha_5 = & +0.0002139607\ldots & \alpha_{13} = & -0.0000000004\ldots \\
\alpha_6 = & -0.0000093400\ldots & \alpha_{14} = & +0.0000000009\ldots \\
\alpha_7 = & -0.0000154019\ldots & \alpha_{15} = & +0.0000000005\ldots.
\end{array}
$$

It seems that α_n is approaching 0 rather rapidly! Of course $\alpha_n \neq 0$ by Lindemann's famous theorem [4].

Szegö [**8**] initiated the study of the zeros of $E_n(z)$. It is convenient to scale down by a factor of n and let

$$\zeta_\nu = \frac{z_\nu}{n}.$$

Szegö proved that the ζ_ν cluster around the simple closed curve $\Gamma = \{z\colon |ze^{1-z}| = 1, |z| \leqslant 1\}$ as $n \to \infty$ and that the proportion which cluster along a given arc of Γ is asymptotic to the change in

$$\frac{1}{2\pi}\arg ze^{1-z}$$

as z varies along the arc. We mention that this implies that the proportion of zeros of E_n with negative real part is asymptotically

$$\frac{1}{2} + \frac{1}{\pi e} = 0.617099\ldots$$

since the arc of Γ which lies in the half plane $\operatorname{Re} z \leqslant 0$ has endpoints $z = \pm i/e$.

Buckholtz [**1**] has shown that the ζ_ν all lie strictly outside Γ and are within a distance $2e/n^{1/2}$ of Γ. By the Eneström-Kakeya theorem on polynomials with monotone coefficients (see Polya-Szegö [7, part III, problem 23]) all the ζ_ν are inside the unit circle $|z| = 1$. Moreover, Newman and Rivlin [**5**], [**6**] have established that the region $y^2 \leqslant cx$ has no zeros z_ν (no scaling) if c is any positive number such that $ce^c < \pi/2$; their paper [**5**] also contains a figure showing the location of the zeros of $E_n(z)$ for $n \leqslant 47$. The regular spacing of the z_ν is quite striking as is the parabolic region free of zeros. Saff and Varga considered the existence in general of a "parabolic" region free of zeros of the sections of power series of entire functions and have conjectured a precise relationship between the "width" of such a region and the order of the entire function. In [3], with Edrei, they prove the conjecture for a class of functions; this work also has an extensive bibliography on this and related problems.

The following indicates another aspect of the interesting geometry of the z_ν.

THEOREM. *If β is any positive number for which $\beta < 1 - \log 2 = 0.3068\ldots$, then*

$$|\alpha_n| \leqslant e^{-\beta n}$$

for all sufficiently large n.

Thus, α_n is an exponentially small sum of terms, most of which are exponentially large as functions of n. As a contrast we mention that it is not difficult to prove that $\sum_{\nu=1}^{n} e^{z_\nu}$ does increase exponentially with n.

The proof of the theorem is not difficult. We write

$$e^z = E_n(z) + R_n(z), \tag{1}$$

where

$$R_n(z) = \sum_{k=n+1}^{\infty} \frac{z^k}{k!}. \tag{2}$$

The idea of the proof is roughly as follows. Let $z_\nu = x_\nu + iy_\nu$ be a zero of $E_n(z)$ and consider $\sum e^{-z_\nu}$. If $x_\nu > n(1 - \log 2)$ then e^{-z_ν} is small. If $x_\nu < n(1 - \log 2)$ we use $e^{-z_\nu} = 1/R_n(z_\nu)$. In this case $|z_\nu|$ is not too large because ζ_ν is near Γ as above. Then $1/R_n(z_\nu)$ can be expanded into an absolutely convergent series of increasing powers of z_ν; the first term is $(n+1)!z_\nu^{-n-1}$. Now using the Lagrange interpolation formula we can show that

$$\sum_{\nu=1}^{n} z_\nu^{-m} = \begin{cases} 1/n! & \text{if } m = n+1 \\ 0 & \text{if } 2 \leqslant m \leqslant n \\ -1 & \text{if } m = 1. \end{cases} \tag{3}$$

Thus the $n+1$ in the definition of α_n arises from $m = n+1$ here. Then we show that the contribution of terms with "large" x_ν and $m \geqslant -1$ is small and similarly for terms with "small" x_ν and $m < -1$. These estimations require a bound for $n!$, a bound for the coefficients in the expansion of $R_n(z)^{-1}$ and the fact that the ζ_ν are near Γ. Note that the points z of Γ for which $x = 1 - \log 2$ satisfy $|z| = 1/2$.

Lemma. *With $R_n(z)$ as above,*

$$\frac{1}{R_n(z)} = \frac{(n+1)!}{z^{n+1}}\left(1 + \sum_{k=1}^{\infty} c_k z^k\right),$$

where

$$|c_k| \leqslant \frac{1}{2}\left(\frac{2}{n+2}\right)^k.$$

The series is absolutely convergent for $|z| < (n+2)/2$.

We will prove this lemma and (3) later. Now we give the proof of the theorem. Fix positive numbers $\beta < \gamma^- < \gamma < \gamma^+ < 1 - \log 2$. Define a partition of $\{1, 2, \ldots, n\}$ into $S \cup L$ by $\nu \in S$ if $x_\nu \leqslant n\gamma$ and $\nu \in L$ if $x_\nu > n\gamma$. Note that by Buckholtz's results,

$$\left|\frac{z_\nu}{n}\right| = e^{\frac{x_\nu}{n} - 1} + o(1) \qquad (1 \leqslant \nu \leqslant n). \tag{4}$$

Trivially

$$|e^{-z_\nu}| \leqslant e^{-n\gamma} \qquad (\nu \in L), \tag{5}$$

and by (4),

$$|z_\nu| \geqslant n e^{\gamma^- - 1} \qquad (\nu \in L) \tag{6}$$

and

$$|z_\nu| \leqslant ne^{\gamma^+ - 1} < \frac{n}{2} \qquad (\nu \in S) \tag{7}$$

for sufficiently large n. From (5) we see that

$$\left| \sum_{\nu \in L} e^{-z_\nu} \right| \leqslant ne^{-n\gamma} \leqslant \frac{1}{2} e^{-n\beta} \tag{8}$$

for large n. By (1), (7), and the lemma,

$$\sum_{\nu \in S} e^{-z_\nu} = \sum_{\nu \in S} 1/R_n(z_\nu) = (n+1)! \sum_{\nu \in S} \sum_{k=0}^{\infty} c_k z_\nu^{k-n-1} \tag{9}$$

where $c_0 = 1$. By (3),

$$\sum_{k=0}^{n} \sum_{\nu \in S} c_k z_\nu^{k-n-1} = \frac{1}{n!} - c_n - \sum_{k=0}^{n} \sum_{\nu \in L} c_k z_\nu^{k-n-1}. \tag{10}$$

Then by (9) and (10)

$$\begin{aligned} \sum_{\nu=1}^{n} e^{-z_\nu} = (n+1) + \sum_{\nu \in L} e^{-z_\nu} - (n+1)! \\ \times \left(c_n + \sum_{\nu \in L} \sum_{k=0}^{n} c_k z_\nu^{k-n-1} - \sum_{\nu \in S} \sum_{k=n+1}^{\infty} c_k z_\nu^{k-n-1} \right). \end{aligned} \tag{11}$$

Now by (6), (7), and the lemma, we can bound

$$\left| c_n + \sum_{\nu \in L} \sum_{k=0}^{n} c_k z_\nu^{k-n-1} - \sum_{\nu \in S} \sum_{k=n+1}^{\infty} c_k z_\nu^{k-n-1} \right|$$

from above by

$$\begin{aligned} &\left(\frac{2}{n}\right)^n + n \sum_{k=0}^{n} \left(\frac{2}{n}\right)^k \left(ne^{\gamma^- - 1}\right)^{k-n-1} + n \sum_{k=n+1}^{\infty} \left(\frac{2}{n}\right)^k \left(ne^{\gamma^+ - 1}\right)^{k-n-1} \\ &\leqslant \left(\frac{2}{n}\right)^n + \frac{n^{-n}}{\left(e^{\gamma^- - 1}\right)^{n+1}} \sum_{k=0}^{\infty} \left(2e^{\gamma^+ - 1}\right)^k \\ &= \left(\frac{2}{n}\right)^n + \frac{n^{-n}}{\left(e^{\gamma^- - 1}\right)^{n+1}} \frac{1}{1 - 2e^{\gamma^+ - 1}} \end{aligned} \tag{12}$$

since $e^{\gamma^+ - 1} < 1/2$. It is easy to show that $(n+1)! \leqslant n^2 (n/e)^n$. Thus by (5), (11), and (12),

$$\left| \sum_{\nu=1}^{n} e^{-z_\nu} - (n+1) \right| \leqslant \frac{1}{2} e^{-n\beta} + n^2 \left(\frac{2}{e}\right)^n + \frac{n^2 e^{-n\gamma}}{e^{\gamma^- - 1}\left(1 - 2e^{\gamma^+ - 1}\right)} \tag{13}$$

if n is sufficiently large. Since $(2/e) < e^{-\beta}$ and $\gamma^- > \beta$, the right-hand side of (13) is $\leqslant e^{-n\beta}$ if n is sufficiently large.

Now we prove (3) and the lemma. For any polynomial $Q(z)$ of degree $\leqslant n-1$,

$$Q(z) = E_n(z) \sum_{\nu=1}^{n} \frac{Q(z_\nu)}{E_n'(z_\nu)(z - z_\nu)}$$

since both sides are polynomials of degree $\leqslant n-1$ which agree at the n points $z_1, \ldots, z_n$. We observe that

$$E_n'(z_\nu) = -\frac{z_\nu^n}{n!}.$$

We obtain the first formula in (3) by taking $Q(z) \equiv 1$, $z = 0$. The second formula follows from the choice $Q(z) = z^k$, $z = 0$ for each of $k = 1, 2, \ldots, n-1$. Finally, the third formula of (3) is a consequence of the fact that the numbers z_ν^{-1} are roots of $z^n + z^{n-1}/1! + \cdots + 1/n!$ so that their sum is -1.

To prove the lemma we write

$$f(z) = \frac{z^{n+1}}{(n+1)!R_n(z)} = \left(1 + (n+1)! \sum_{k=1}^{\infty} \frac{z^k}{(n+1+k)!}\right)^{-1}$$

and expand the right-hand side as a geometric series. This is legitimate if $|z| < (n+2)/2$, since then

$$\left|(n+1)! \sum_{k=1}^{\infty} \frac{z^k}{(n+1+k)!}\right| < \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^k = 1.$$

The power series we obtain for $f(z)$ is majorized by

$$1 + \sum_{l=1}^{\infty} \left[(n+1)! \sum_{k=1}^{\infty} \frac{z^k}{(n+1+k)!}\right]^l,$$

which in turn is majorized by

$$1 + \sum_{l=1}^{\infty} \left(\sum_{k=1}^{\infty} \frac{z^k}{(n+2)^k}\right)^l.$$

Since

$$1 + \sum_{l=1}^{\infty} \left(\sum_{k=1}^{\infty} w^k\right)^l = 1 + \sum_{l=1}^{\infty} \left(\frac{w}{1-w}\right)^l = \frac{1-w}{1-2w} = 1 + \frac{w}{1-2w}$$
$$= 1 + \frac{1}{2} \sum_{k=1}^{\infty} 2^k w^k,$$

the lemma follows.

We remark that $R_n(z)$ has no zeros z with $|z| \leqslant n+2$. This fact can be proved exactly as the Eneström-Kakeya theorem mentioned earlier. Thus the series for $z^{n+1}/R_n(z)$ actually converges absolutely for $|z| \leqslant n+2$. Since the z_ν satisfy $|z_\nu| \leqslant n$ we have the formulae

$$\sum_{\nu=1}^{n} e^{-z_\nu} = \sum_{\nu=1}^{n} 1/R_n(z_\nu) = n + 1 - (n+1)!c_n + (n+1)! \sum_{k=0}^{\infty} c_{n+1+k} \sum_{\nu=1}^{n} z_\nu^k \quad (14)$$

and

$$\limsup_{m \to \infty} |c_m|^{1/m} < \frac{1}{n+2}. \quad (15)$$

Thus, it is possibly the case that α_n is asymptotic to $-(n+1)!\, c_n$ but it is not clear how to better estimate c_k for $k \leqslant 2n$ and so prove this.

REFERENCES

1. J. D. Buckholtz, A characterization of the exponential series, *Amer. Math. Monthly*, 73 (1966) 121–123.
2. J. B. Conrey and A. Ghosh, Zeros of derivatives of the Riemann zeta function near the critical line, in preparation.
3. A. Edrei, E. B. Saff, and R. S. Varga, Zeros of Sections of Power Series, Lecture Notes in Mathematics 1002, Springer, Berlin, 1983.
4. F. Lindemann, Uber die Zahl π, *Math Annalen*, 20 (1882) 213–225.
5. D. J. Newman and T. J. Rivlin, The zeros of the partial sums of the exponential function, *J. Approx. Theory*, 5 (1972) 405–412.
6. D. J. Newman and T. J. Rivlin, Correction: The zeros of the partial sums of the exponential function, *J. Approx. Theory*, 16 (1976) 229–300.
7. G. Polya and G. Szegö, Problems and Theorems in Analysis I, Springer-Verlag, New York 1972.
8. G. Szegö, Uber eine Eigenschaft der Exponentialreihe, *Sitzungsber. Berl. Math. Ges.*, 21 (1922) 59–64.

Letters to the Editor

Editor:

The method used in Norwegian books (e.g., [**1**]) to prove the reflective property of a parabola seems more direct than that used by Robert Williams [**2**]. Referring to Williams's diagram, let Q by the y-intercept of l_2. Then, since l_2's slope is $x_0/2c$ and the parabola's equation is $y = x^2/4c$, $Q = (0, -y_0)$. It follows that $QFPD$ is a parallellogram (where $D = (x_0, -c)$), whence its diagonal QP bisects $\angle FPD$ and $\alpha = \beta$.

REFERENCES

1. A. Søgaard and R. T. Lyche, *Mathematikk III for Realgymnaset*, Glydendal Norsk Forlag, Oslo, 1948.
2. R. Williams, A proof of the refective property of the parabola, this MONTHLY, 94 (1987) 667–68.

Ragnar Dybvik
N-6630 Tingvoll
Norway

Editor:

With reference to your recent editorial 'Strings, Substrings and the Nearest Integer Function' (*Amer. Math. Monthly*, 94 (Nov. 87) 855–860), I note that in yóur second example you express the nth Fibonacci number F_n using the nearest integer function. It may be of interest to note that F_n can be expressed recursively using the nearest integer function; *viz.*, F_n is the nearest integer to the geometric mean of F_{n-1} and F_{n+1}. (See my Lemma 1 on diophantine defining relationals in *Abstracts AMS*, 8 (Oct. 87) 437–438.) Indeed, there exist infinitely many recursions of this type; e.g., F_n is the nearest integer to the geometric mean of F_{n-2}, F_{n-1}, and F_{n+2}.

Albert A. Mullin
506 Seaborn Drive
Huntsville, AL 35806

Editor:

I read with interest your paper in the November 1987 MONTHLY about strings. Concerning your quest for "a legal English word that has initial and final proper substrings of length 5 or more that are the same," here are some suggestions.

1. UNDERFUNDER. With all the complaints we hear nowadays about people and institutions being underfunded, surely the responsibility lies with an underfunder or two. If you do not like this word, you certainly won't enjoy
2. OVERSHOVER. Haven't we all been asked occasionally to shove over? The person making the request is an overshover.
3. UPPERSUPPER. Meal served to people in the upper berths of combination dining/sleeping cars, perhaps.
4. INTERSPLINTER. I give up, I can't think of a remotely plausible definition for this one.

Gerald Myerson
Macquarie University
New South Wales
Australia

[Dr. Jim Saxe supplies "UNDERGROUNDER", which is actually in the *Oxford English Dictionary*.—Ed.]

Editor:

Shanks' Identity is Elementary

The identity in question, discussed in [**1**], is

$$\sqrt{5} + \sqrt{(22 + 2\sqrt{5})} = \sqrt{(11 + 2\sqrt{29})} + \sqrt{(16 - 2\sqrt{29} + 2\sqrt{\{55 + 10\sqrt{29}\}})}.$$

An elementary proof of the identity follows from the observation that for positive x, y the quantities

$$A = xy + \sqrt{\{x^2 + 2xy + y^2\}},$$
$$B = y + \sqrt{\{x^2(1 + y^2) + 2xy \cdot x\}}$$

are both disguised forms of $x + y + xy$. Taking $x, y = \sqrt{(11 \pm 2\sqrt{29})}$—$x$ with minus, y with plus—we find that A reduces to the left member of the identity and B to the right.

I discovered this in the fall of 1984 and learned six months later of a similar discovery by R. Bundy. The identity has been subject to a complicated treatment in a recent article [**2**].

REFERENCES

1. D. Shanks, Incredible identities, *Fibonacci Quart.*, 12 (1974) 271.
2. C. E. van der Ploeg, Duality in nonnormal quartic fields, this MONTHLY, 94 (1987) 279–284.

J. G. Wendel
The University of Michigan,
Ann Arbor, MI 48109

Editor:

In his Note "Pointwise Limits of Analytic Functions" (this MONTHLY, 90 (1983) 391–394) Kenneth R. Davidson reviews the classical theorems needed to prove the basic result that if f_n is a sequence of analytic functions on an open subset Ω of the plane which converges pointwise to f on Ω then there is a dense open subset Ω_0 of Ω such that the convergence is uniform on compact sets (u.c.c.), and f is analytic, in Ω_0. On the relatively closed nowhere dense set $C = \Omega \setminus \Omega_0$, f is merely the pointwise limit of continuous functions (that is, in Baire Class 1, or BC1). In the concluding section "What Limits Are Possible" he gives an argument purporting to prove that given any such sets $\Omega, \Omega_0, C = \Omega \setminus \Omega_0$ and any f analytic on Ω_0 and BC1 on C, then $f = \lim f_n$ for some analytic function sequence f_n on Ω with u.c.c. convergence on Ω_0. Unfortunately the truth is much more complex than this, and, despite the statement in the opening paragraph that, "The results in this paper... do not appear to have been collected together before," there is a literature going back to 1901 and culminating in M. Lavrentiev's comprehensive treatment of the topic in **[8]** of 1936.

W. F. Osgood **[1]** proved that if analytic $f_n \to f$ in a region Ω then f is analytic on a dense open subset of Ω and showed "by methods due to Runge, Acta Mathematica 6 (1885) 229" that "the regions of analyticity may be infinite." P. Montel **[2]** gave a construction to show that f could be continuous yet fail to be analytic on the whole of Ω. Later in **[3]** (Chapter V, pp. 108–122) Montel proved the u.c.c. convergence of f_n to f on a dense open subset Ω_0 of Ω and asked about the structure of $C = \Omega \setminus \Omega_0$ and whether any f analytic on Ω_0 and BC1 on C could be represented as the limit of a sequence f_n of analytic functions on Ω. Here Ω was assumed to be a simply connected domain and Montel established the simple necessary condition on C that the components of Ω_0 be simply connected. Indeed, if a contour $\Gamma \subseteq \Omega_0$ and $z \in$ interior Γ, then

$$f(z) = \lim f_n(z) = \lim \frac{1}{2\pi i} \int_\Gamma \frac{f_n(w)}{(w-z)}\, dw$$

uniformly in z bounded away from Γ, by dominated convergence of f_n on Γ, whence interior $\Gamma \subseteq \Omega_0$. Montel's problem, as it came to be called, was finally solved in papers of F. Hartogs & A. Rosenthal **[6, 7]** and M. Lavrentiev **[4, 5, 8]**. Extensions to the analogous problem of sequences of harmonic functions in 3 real variables were given by M. Keldysch & M. Lavrentiev **[9, 10]**.

The necessary and sufficient conditions on C and f are rather complicated and the reader is referred to the literature. For some C Davidson's argument does work and f can be taken arbitrary analytic on $\Omega \setminus C$ and arbitrary BC1 on C. For other C there is a more restricted class of limit functions f which have the characteristic property that the values of $f|_C$ partly determine the values of $f|_\Omega$. The following counterexample (not in the literature) shows that even if Montel's necessary condition is satisfied there are cases where C does not allow any limit function f at all.

THEOREM. *Let Ω be a simply connected domain and C a relatively closed nowhere dense subset of Ω such that $\Omega_0 = \Omega \setminus C$ is a disjoint union of open discs. Then there does not exist any function f on Ω such that $f = \lim f_n$ for a sequence of analytic functions f_n on Ω with Ω_0 as the set of u.c.c. convergence.*

Proof. Suppose an f and sequence f_n did exist. Then $f_n|C$ is locally bounded on a dense open subset of C (as in Davidson's Proposition 1, p. 393) so there is an open disc $D = D(z_0, r) \subseteq \Omega$ of radius r and centre $z_0 \in C$, and an $M > 0$ such that $|f_n(z)| \leqslant M$ for all n and $z \in D \cap C$. Let $D^* = D(z_0, \frac{1}{2}r)$.

Any disc S of Ω_0 of radius $\leqslant \frac{1}{4}r$ which meets D^* has $|f_n| \leqslant M$ on $\partial S \in D \cap C$ so that $|f_n| \leqslant M$ on $S \cup \partial S$ by the maximum modulus principle. Such discs S exist because any point of C, in particular z_0, is a limit of arbitrarily small discs of Ω_0. At most a finite number $S_1, \dots, S_k$ of discs of Ω_0 of radius $> \frac{1}{4}r$ meet D^* and thus $|f_n| \leqslant M$ in $R = D^* \setminus S_1 \setminus S_2 \setminus \cdots \setminus S_k$.

It follows that int $R \subseteq \Omega_0$ (e.g., by Davidson's Proposition 3, p. 393), but it is clearly geometrically impossible for int R to be a disjoint union of open discs or parts of such discs intersected with D^*. So after all f and f_n do not exist. Q.E.D.

REFERENCES

1. W. F. Osgood, Note on the functions defined by infinite series whose terms are analytic functions of a complex variable, *Ann. of Math.*, (2)3 (1901) 25–34.
2. P. Montel, Sur les séries de fonctions analytiques, *Bull. Sc. Math.*, 30 (1906) 189–192.
3. ______, Leçons sur les séries de pôlynomes à une variable complexe, Paris 1910.
4. M. Lavrentiev, Sur un problème de M. P. Montel, *Comptes Rendus Acad. Sci.*, 184 (1927) 1634–35.
5. ______, Sur un problème de M. P. Montel, Ibid., 188 (1929) 689–691.
6. F. Hartogs and A. Rosenthal, Über Folgen analytischer Funktionen, *Math. Annalen*, 100 (1928) 212–263.
7. ______, Ibid., 104 (1931) 606–610.
8. M. Lavrentiev, Sur les fonctions d'une variable complexe représentable par des séries de pôlynomes, Actualités Scientifiques et Industrielles No. 441, Paris 1936.
9. M. Keldysch and M. Lavrentiev, Sur les suites de pôlynomes harmoniques, *Comptes Rendus Acad. Sci.*, 202 (1936) 1149–51.
10. ______, Sur les suites convergentes de pôlynomes harmoniques, *Trav. Inst. Math. Tbilissi*, 1 (1937) 165–184.

M. J. Pelling
Department of Mathematics
University College
London WC1E 6BT
England

NOTES

EDITED BY DENNIS DETURCK, DAVID J. HALLENBECK, AND ANITA E. SOLOW

On Continued Fractions and a Certain Example of a Sequence of Continuous Functions

J. FABRYKOWSKI*
Department of Mathematics, University of Alberta, Edmonton, Alberta, Canada T6G 2G1

In this note we shall construct an interesting example of a sequence of continuous functions $f_n(x) \in C[0,1]$ such that

$$\lim_{n\to\infty} f_n(x) = \begin{cases} +\infty & \text{if } x \text{ is irrational} \\ O(1) & \text{if } x \text{ is rational.} \end{cases} \qquad (*)$$

We require some basic facts from the theory of continued fractions, which we will not prove here. There is an extensive literature where one can find more details (see, e.g. **[1]**, **[2]**, **[3]**).

Define a finite continued fraction as any expression of the form

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \begin{matrix} \\ \vdots \\ \end{matrix} + \cfrac{1}{a_n}}}},$$

where a_i's are real numbers and $a_i \neq 0$ for $1 \leqslant i \leqslant n$. For case or reference we will designate the above continued fraction by the symbol

$$\langle a_0, a_1, \dots, a_n \rangle. \qquad (1)$$

The numbers $a_0, a_1, a_2, \dots$ are called the quotients of the continued fraction (1) and any expression $C_t = \langle a_0, a_1, \dots, a_t \rangle$, $0 \leqslant t \leqslant n$ is said to be the tth convergent to the continued fraction. If all quotients are integers and moreover $a_i > 0$ for $i \geqslant 1$, then (1) is called a simple finite continued fraction.

We have the following:

THEOREM 1. *Let* (1) *be a simple finite continued fraction. Let the two finite sequences of real numbers p_t and q_t be defined inductively by*:

$$p_0 = a_0, \quad p_1 = a_1 a_0 + 1, \quad p_t = a_t p_{t-1} + p_{t-2}, \qquad 2 \leqslant t \leqslant n, \qquad (2)$$

*Supported by NSERC Grant #A3062.

and

$$q_0 = 1, \quad q_1 = a_1, \quad q_t = a_t q_{t-1} + q_{t-2}, \qquad 2 \leqslant t \leqslant n. \tag{3}$$

Then the following facts hold: The p_i's and q_i's are integers, the q_i's are positive and

$$C_t = \frac{p_t}{q_t} \quad \textit{for } 0 \leqslant t \leqslant n. \tag{4}$$

$$q_t \geqslant t, \textit{ with inequality when } t > 3. \tag{5}$$

The sequence of even convergents C_{2t} is strictly increasing with t ($t \geqslant 0$), while the odd convergents C_{2t+1} strictly decrease. Also every odd convergent is greater than any even convergent, that is,

$$C_0 < C_2 < C_4 < \cdots < C_5 < C_3 < C_1. \tag{6}$$

THEOREM 2. *Let $a_0, a_1, a_2, \ldots,$ be an infinite sequence of integers, all positive except possibly a_0. If $C_n = \langle a_0, a_1, a_2, \ldots, a_n \rangle$, then $\lim_{n \to \infty} C_n$ exists and is finite.*

Theorem 2 justifies the following definition:

DEFINITION. Let $a_0, a_1, a_2, \ldots$ be a sequence of integers, all positive except possibly a_0. The expression $\langle a_0, a_1, a_2, \ldots \rangle$ is called an infinite simple continued fraction and is defined to be equal to the number $\lim_{n \to \infty} C_n$, where $C_n = \langle a_0, a_1, \ldots, a_n \rangle$.

It may be proved that Theorem 1 holds as well for a simple infinite continued fraction. Consider an irrational number x. We now show how to find representation of x as a simple infinite continued fraction. We write $x = a_0 + \xi_0$, where $a_0 = [x]$ – the greatest integer $\leqslant x$ and $0 < \xi_0 < 1$. Let $1/\xi_0 = a_1'$, so $x = a_0 + 1/a_1'$. Define $a_1 = [a_1']$, thus $a_1' = a_1 + \xi_1$, where $0 < \xi_1 < 1$. Similarly, let $1/\xi_1 = a_2'$ and define $a_2 = [a_2']$, so $a_2' = a_2 + \xi_2$ and $x = a_0 + 1/(a_1 + (1/a_2'))$. Continuing this procedure, we obtain an infinite sequence at integers: $a_0, a_1, a_2, \ldots$ with $a_i \geqslant 1$ for $i \geqslant 1$ and an infinite sequence of real numbers a_i' ($i \geqslant 1$) such that

$$x = \langle a_0, a_1' \rangle = \langle a_0, a_1, a_2' \rangle = \cdots = \langle a_0, a_1, \ldots, a_n, a_{n+1}' \rangle = \cdots.$$

It is well known that if $C_n = \langle a_0, a_1, \ldots, a_n \rangle$, where $a_0, a_1, \ldots, a_n$ are found by means of the above algorithm, then $x = \lim_{n \to \infty} C_n$ and that the representation of x as an infinite simple continued fraction is unique. The convergents to any irrational number have a certain characteristic property, described in the following:

THEOREM 3. *Let x be an irrational number and a/b rational with $(a, b) = 1$ and $b > 0$. If*

$$\left| x - \frac{a}{b} \right| < \frac{1}{2b^2},$$

then a/b is one of the convergents to x.

Now we are in a position to construct the sequence of continuous functions $f_n(x)$ satisfying ($*$).

Let $r_1, r_2, r_3, \ldots$ be enumeration of the rationals in $[0, 1]$ in the following manner:

$$0, 1, \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{3}{4}, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, \frac{1}{6}, \frac{5}{6}, \frac{1}{7}, \ldots.$$

We shall show that the sequence

$$f_n(x) = \sum_{k=1}^{n} \left\{ \prod_{j=1}^{k} 2j|x - r_j| \right\}^{1/k}, \qquad n = 1, 2, 3, \ldots$$

has the desired property. It is obvious that for every $n \geqslant 1$, $f_n(x) \in C([0,1])$. For any rational x there exists a number $n_0 = n_0(x)$ that $f_n(x) = f_{n_0}(x)$ for $n \geqslant n_0$; therefore, for rationals $x, \lim_{n\to\infty} f_n(x) = f_{n_0}(x)$, thus it is finite.

Suppose now that x is irrational. By enumeration of r_j it follows that if we write $r_j = u_j/v_j$, where u_j and v_j are positive integers such that $(u_j, v_j) = 1$, then obviously $v_j < j$ and, therefore, by (5) none of r_j is a convergent to x. From Theorem 3 it follows that

$$|x - r_j| = \left| x - \frac{u_j}{v_j} \right| \geqslant \frac{1}{2v_j^2} > \frac{1}{2j^2}. \tag{7}$$

We have the obvious inequality

$$\left\{ \prod_{j=1}^{k} 2j|x - r_j| \right\}^{1/k} > \min_{1 \leqslant j \leqslant k} 2j|x - r_j|,$$

so by (7)

$$\left\{ \prod_{j=1}^{k} 2j|x - r_j| \right\}^{1/k} > 2j\frac{1}{2j^2} = \frac{1}{j} \geqslant \frac{1}{k}. \tag{8}$$

Therefore, if x is irrational,

$$f_n(x) > \sum_{k=1}^{n} \frac{1}{k}, \quad \text{thus} \lim_{n\to\infty} f_n(x) = \infty.$$

Related to our example we mention a theorem that goes back to W. F. Osgood (see [4]).

THEOREM. *If a function $f(x)$ can be represented as a limit of an everywhere convergent sequence $f_n(x)$ of continuous functions, then $f(x)$ is continuous except at a set of points of first category.*

Since the set of rational numbers is of first category and the set of irrational numbers is then of second category (see [4]), therefore, it is not possible to have an example of a sequence $f_n(x)$ of continuous functions such that

$$\lim_{n\to\infty} f_n(x) = \begin{cases} +\infty & \text{if } x \text{ is rational} \\ O(1) & \text{if } x \text{ is irrational} \end{cases}.$$

Another example of a sequence of continuous functions $f_n(x)$ satisfying property $(*)$ was found by Professor A. Meir.

REFERENCES

1. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, Oxford, 1960.
2. C. D. Olds, Continued Fractions, The New Mathematical Library, MAA, 1963.
3. J. E. Shockley, Introduction to Number Theory, Holt, Rinehart and Winston, 1967.
4. J. C. Oxtoby, Measure and Category, Springer-Verlag, 1971.

Using Pythagorean Triangles to Approximate Angles

W. S. ANGLIN
Department of Mathematics and Statistics, McGill University, Montreal, PQ, Canada H3A 2K6

A Pythagorean triangle is a right triangle with integer sides. A well-known example is the triangle with sides 3, 4, and 5. Unfortunately, the angles of these triangles are not as "nice" as the sides: whether expressed in degrees or radians, they are always irrational.

If a Pythagorean triangle has an angle of A degrees and $|A - 20| < 1/100$, we shall say that the triangle "approximates 20 degrees to within one hundredth." Let B be any given angle in degrees with $0 < B < 90$. Let e be any real number such that $0 < e < 1$, $e < B$, and $e < 90 - B$. The question we answer in this note is: what is the smallest Pythagorean triangle which approximates B degrees to within e?

We begin by defining "smaller." If (a, b, c) and (a', b', c') are two Pythagorean triangles with sides $a < b < c$ and sides $a' < b' < c'$, then (a, b, c) is "smaller" than (a', b', c') just in case (i) $c < c'$ or (ii) $c = c'$ and $b < b'$. Roughly speaking, this means we order the triangles by hypotenuse. For example, the triangle $(3, 4, 5)$ is the smallest of all Pythagorean triangles since it has the smallest hypotenuse.

Pythagorean triangles whose sides are relatively prime are called "primitive." For the purpose of finding a *smallest* triangle approximating a given angle, we need only consider primitive triangles. We thus confine our attention to them. Since exactly one of the two smaller sides of a primitive Pythagorean triangle is even, the following well-known theorem gives a complete characterization of these triangles.

THEOREM 1. *$a^2 + b^2 = c^2$ with a an even integer and a, b, c relatively prime integers iff for some relatively prime positive integers u and v with $v < u$ and u, v not both odd,*

$$a = 2uv, \qquad b = u^2 - v^2, \qquad \text{and} \quad c = u^2 + v^2.$$

For a proof the reader may consult [**1**, pp. 138–139].

Roughly speaking, we shall approximate a given angle B by means of approximating its tangent or cotangent. In this connection we need the following definitions.

$$X = \tan(B - e) + \sec(B - e)$$
$$Y = \tan(B + e) + \sec(B + e)$$
$$X' = \cot(B + e) + \csc(B + e)$$
$$Y' = \cot(B - e) + \csc(B - e).$$

We now have the "Approximation Theorem."

THEOREM 2. *Let u and v be positive integers. Then $X < u/v < Y$ iff $(2uv, u^2 - v^2, u^2 + v^2)$ approximates B to within e by means of the angle opposite $u^2 - v^2$ and $X' < u/v < Y'$ iff $(2uv, u^2 - v^2, u^2 + v^2)$ approximates B to within e by means of the angle opposite $2uv$.*

Proof. Suppose $X < u/v < Y$. The given conditions on B and e imply that $1 < X$ and, hence, $v < u$ (since $X = (\sin(B - e) + 1)/\cos(B - e)$). Since $X < u/v < Y$, it follows that $-1/X < -v/u < -1/Y$. Adding these inequalities, we obtain $X - 1/X < (u^2 - v^2)/uv < Y - 1/Y$. Moreover, as a straightforward calculation will show, $X - 1/X = 2\tan(B - e)$ and $Y - 1/Y = 2\tan(B + e)$. Hence, we obtain $\tan(B - e) < (u^2 - v^2)/2uv < \tan(B + e)$. Hence, in $(2uv, u^2 - v^2, u^2 + v^2)$ the angle opposite $u^2 - v^2$ is between $B - e$ and $B + e$ degrees.

Suppose now that the right-hand side of the first equivalence holds. Then where $z = u/v$ we have $2\tan(B - e) < z - 1/z < 2\tan(B + e)$. Hence, $z^2 - (2\tan(B + e))z - 1 < 0$ and $0 < z^2 - (2\tan(B - e))z - 1$. The first of these inequalities implies that z is between the two roots of $z^2 - (2\tan(B + e))z - 1$. Thus $z < \tan(B + e) + \sqrt{\tan^2(B + e) + 1} = \tan(B + e) + \sec(B + e) = Y$. Similarly, from $0 < z^2 - (2\tan(B - e))z - 1$ it follows that $X < z$.

The proof for the second equivalence is similar.

Finally, we have the "Ordering Theorem."

THEOREM 3. *Let V be the smallest positive integer such that $[XV] + 1 < YV$ and let $U = [XV] + 1$. Then $(2UV, U^2 - V^2, U^2 + V^2)$ is the smallest Pythagorean triangle approximating B degrees to within e by means of the angle opposite the odd side.*

Moreover, the corresponding result holds for the "even side case" (with X', Y' replacing X, Y, respectively).

Proof. Since $X < U/V < Y$, the above triangle does approximate B degrees to within e by means of the angle opposite $U^2 - V^2$.

Let $T = (2u'v', u'^2 - v'^2, u'^2 + v'^2)$ be the smallest such triangle. Then $X < u'/v' < Y$ and, hence, $[Xv'] + 1 < Yv'$. Thus $V \leqslant v'$. If $U < u'$, then $U^2 + V^2 < u'^2 + v'^2$ against T's minimality. Hence, $u' \leqslant U$. Since $Xv' < u'$, it follows that $U = [XV] + 1 \leqslant [Xv'] + 1 \leqslant u'$. Thus $u' = U$. By T's minimality, $v' \leqslant V$ and, hence, $v' = V$. Thus the two triangles are the same.

The proof for the corresponding "even side result" is similar.

For example, suppose we want the smallest Pythagorean triangle approximating 20 degrees to within 1/100. We compute $X = 1.42788278$, $Y = 1.42841330$, $X' = 5.66838919$ and $Y' = 5.67417730$. By straightforward testing (on a programmable calculator) we find that $V = 208$ is the smallest positive integer such that $[XV] + 1 < YV$ and $V' = 46$ is the smallest positive integer such that $[X'V'] + 1 < Y'V'$. Let $U = [XV] + 1 = 297$ and let $U' = [X'V'] + 1 = 261$. The smallest triangle to approximate 20 degrees to within 1/100 by means of the angle opposite the odd side is $(2UV,\ U^2 - V^2,\ U^2 + V^2) = (123552,\ 44945,\ 131473)$ and the smallest triangle to approximate 20 degrees to within 1/100 by means of the angle opposite the even side is $(2U'V',\ U'^2 - V'^2,\ U'^2 + V'^2) = (24012, 66005, 70237)$. The second is smaller.

I thank Professor J. Lambek for his kind help and encouragement.

REFERENCES

1. Ivan Niven and H. S. Zuckerman, An Introduction to the Theory of Numbers, 4th ed., John Wiley & Sons, New York, 1980.

Convergence-Preserving Functions

GERALD WILDENBERG*
Dept. of Mathematics, St. John Fisher College, Rochester, NY 14618

It is common for calculus texts to include as an exercise the question of whether the convergence of Σa_n (with a_n real) implies that Σa_n^2 also converges. If Σa_n is absolutely convergent then the comparison test applied to the tails of Σa_n and Σa_n^2 will show that Σa_n^2 is convergent. But if we do not have absolute convergence it is easy to come up with a counter-example (e.g., $a_n = (-1)^n/\sqrt{n}$).

The situation is more interesting if we replace a_n^2 with $f(a_n)$ and then ask: Does the convergence of Σa_n imply the convergence of $\Sigma f(a_n)$? If f is x^3 or arcsin x, or $x/(1+|x|)$, we have questions that will make most mathematicians pause. In this note we answer, negatively, the question as to whether there are any nontrivial real functions for which Σa_n converges only if $\Sigma f(a_n)$ converges.

DEFINITION. A function on the reals, f, is *convergence preserving* (abbreviated CP) if for every convergent series Σa_n, the series $\Sigma f(a_n)$ also converges. For example, a constant multiple function, $f(x) = kx$ is CP. As we shall show, there are essentially no other CP functions.

The following properties are immediate.

(1) If f is CP, then $f(0) = 0$. (Consider the 0 sequence.)

(2) If f and g are CP, so are $f+g$, kf, $f\circ g$ and, therefore, h, where $h(x) = f(kx)$. (Here k is a constant.)

LEMMA 1. *If f is CP, then there exists a real number M and an $\varepsilon > 0$ such that*

$$f(x)/x < M \quad \text{for } 0 < x < \varepsilon.$$

Proof. If not, then for each n we can find x_n with

$$0 < x_n < 1/n^2 \quad \text{and} \quad f(x_n)/x_n > n.$$

Define j_n to be the least integer that is $\geqslant 1/(n^2 x_n)$. Then clearly $j_n x_n < 2/n^2$. Now consider the series

$$\underbrace{x_1 + x_1 + \cdots + x_1}_{j_1 \text{ times}} + \underbrace{x_2 + x_2 + \cdots\ x_2}_{j_2 \text{ times}} + \cdots.$$

This converges by comparison with $\Sigma 2/n^2$.

When we apply f to each term of the series and sum the first n blocks we get

$$\sum j_n f(x_n) \geqslant \sum n j_n x_n \geqslant \sum 1/n,$$

which diverges, contradicting that f is convergence preserving.

LEMMA 2. *If f is CP, then f is differentiable at* 0.

Proof. Consider D^+f, D_+f, D^-f, D_-f, the upper and lower derivatives of f from the right and left at 0. (So, for example, D_+f is the lower-right derivative of f at 0, that is, the lim inf of $f(x)/x$ as $x \to 0^+$.) We claim that $D_-f \geqslant D^+f$.

*The author thanks John Kennison for his suggestions regarding this paper.

Suppose instead that $D_-f < s < t < D^+f < \infty$. Then let $\{d_n\}$ be a monotonic decreasing sequence with $0 < d_n < 2^{-n}$, but $f(d_n)/d_n > t$ for all n.

Similarly, let $\{e_n\}$ be a monotonic increasing sequence with

$$-2^{-n} < e_n < 0 \quad \text{and} \quad f(e_n)/e_n < s \quad \text{for all } n.$$

We can further assume that

$$d_1 > |e_1| > d_2 > |e_2| > \cdots.$$

We now construct a series whose underlying sequence is

$$\{a_i\} = d_1, \dots, d_1, e_1, \dots, e_1, d_2, \dots, d_2, e_2, \dots .$$

The number of d_i in each block, say j_i, is defined by letting j_i be the smallest integer for which $j_i d_i > 1/i$. Similarly, k_i is the smallest integer such that $k_i e_i < -1/i$, where k_i is the number of e_i's. The series Σa_i is convergent, but compare $\Sigma f(a_i)$ to

$$\sum (tj_n d_n + sk_n e_n) = \sum (t-s) j_n d_n + \sum s(j_n d_n + k_n e_n).$$

(Note that $\Sigma(t-s)j_n d_n$ diverges and $s(\Sigma j_n d_n + k_n e_n)$ converges.) We see that $\Sigma f(a_i)$ is divergent. From this contradiction we conclude that $D_-f \geqslant D^+f$. Similarly, considering $f(-x)$, $D_+f \geqslant D^-f$. Therefore,

$$D_+f \geqslant D^-f \geqslant D_-f \geqslant D^+f \geqslant D_+f.$$

So f is differentiable at 0.

LEMMA 3. *Suppose that $f(0) = f'(0) = 0$. Let A, B, $\varepsilon > 0$ be given. Then there exists $e > 0$ and a positive integer k such that*

$$|e| < \varepsilon \quad \textit{and} \quad A < k|e| < A + \varepsilon \quad \textit{and} \quad k|f(e)| < B.$$

There also exists $e < 0$ with the same properties.

Proof. We may as well assume that $\varepsilon < 1$. Choose e with $|f(e)|/e < B/(A+1)$ and $0 < e < \varepsilon$. Let k be the smallest integer with $k > A/e$. Then

$$k|f(e)| < Bke/(A+1) < B(A+\varepsilon)/(A+1) < B.$$

The case with $e < 0$ is essentially the same.

THEOREM. *f is CP iff there exists a constant k with $f(x) = kx$ in a neighborhood of 0.*

Proof. Assume that f is CP. We may as well assume that $f'(0) = 0$, otherwise we could replace $f(x)$ by $f(x) - kx$, where $k = f'(0)$. Using the indirect method, we further assume that $f(x)$ is not identically zero in any neighborhood of 0. Then we can obtain a sequence $\{d_n\}$ with

$$0 < f(d_n) < d_n < 2^{-n}.$$

[Note that $|f(x)|$ is less than $|x|$ near 0 as $f'(0) = 0$. We can require d_n and $f(d_n)$ to be positive, otherwise $f(x)$ can be replaced by $f(-x)$ or $-f(x)$ or $-f(-x)$.]

Let j_n be the least integer for which $j_n f(d_n) > 1/n$. From Lemma 3, there exists e_n with $-2^{-n} < e_n < 0$ and k_n with $j_n d_n < k_n|e_n| < j_n d_n + 2^{-n}$ and $k_n|f(e_n)| < 1/n^2$.

Consider now a series whose underlying sequence consists of d_i and e_i, interleafed where j_n and k_n are the number of terms in the nth block of d_i and e_i,

respectively. For each i we interleaf the d_i and e_i so that the partial sums within the ith block of d_i and e_i do not exceed 2^{-n+1} in absolute value. Since the ith block sums to $j_i d_i - k_i|e_i|$, the series is convergent. Applying f we get a series where the ith block sums to $j_i f(d_i) + k_i f(e_i)$. This series diverges, since $k_i f(e_i)$ is convergent but $j_i f(d_i)$ is not. This contradicts the CP property for f, proving the proposition. (The converse is, of course, obvious.)

Remark. While the main theorem is apparently a negative result, it does guarantee the existence of many series having strange properties. For example, there is a series $\{a_i\}$ for which Σa_i^3 diverges but $\Sigma \tan a_i$ converges, otherwise $(\arctan x)^3$ is CP. Similarly, there is a $\{b_i\}$ with Σb_i^3 convergent and $\Sigma \tan b_i$ divergent, else $\tan \sqrt[3]{x}$ is CP.

The Diophantine Equation $x^2 + y^m = z^{2n}$

DAVID W. BOYD
Department of Mathematics, University of British Columbia, Vancouver, Canada V6T 1Y4

We will show here that the equation of the title has infinitely many primitive solutions in integers provided m and n are relatively prime. In fact, our proof gives a procedure for finding all solutions.

The result requires little more than unique factorization and thus could be presented at an early stage in a course in elementary number theory. It would be enough to present a couple of special cases, say $x^2 + y^3 = z^4$ or $x^2 + y^4 = z^6$. Of course, one would have first presented the classical equation $x^2 + y^2 = z^2$, which is also a special case.

The equation $x^2 + y^3 = z^4$ has been considered by Mathieu [**1**, p. 664] and Kiss [**2**]. The former exhibited an infinite family of solutions by using the Pell equation $a^2 - 2b^2 = 1$. The latter found all solutions with x, y, and z relatively prime. One of the cases he considered required factorization in the field $Q(\sqrt[3]{2})$. Thus the tools used in these papers are a little more sophisticated than those needed here. In addition, the methods do not seem capable of treating the more general equation we consider.

The word *primitive* used in the first paragraph requires clarification. We may restrict our attention to solutions with $xyz \neq 0$, $x > 0$, and $z > 0$. Note that if (x, y, z) solves $x^2 + y^m = z^{2n}$, then so does $(k^{mn}x, k^{2n}y, k^m z)$ for any integer $k > 0$. A *primitive solution* (x_0, y_0, z_0) is one which is not of the form $(k^{mn}x, k^{2n}y, k^m z)$ for any $k > 1$.

Each solution is thus of the form $(k^{mn}x_0, k^{2n}y_0, k^m z_0)$ for some primitive solution (x_0, y_0, z_0) and some $k \geqslant 1$. We say that the two solutions lie in the same class if they correspond to the same primitive solution. We are really interested in classes of solutions; the primitive solution is the representative of the class which has z minimal.

Note that if $(m, n) \neq (2, 1)$, the equation $x^2 + y^m = z^{2n}$ is not homogeneous so a primitive solution need not have x, y, z relatively prime. For example, $(28, 8, 6)$ is a primitive solution of $x^2 + y^3 = z^4$. Thus the solutions found by Kiss are not all the primitive solutions of this equation, as he observed. We will assume that $m > 1$ and $n > 1$ since the cases $m = 1$ or $n = 1$ are trivial.

Our method is based on the simple observation that, given $m > 1$, any integer $N > 1$ can be written uniquely in the form $N = s_1 s_2^2 \cdots s_{m-1}^{m-1} u^m$, where $s_1, \ldots, s_{m-1}$ are positive, square-free, and pairwise relatively prime. (Here we consider 1 to be square-free and allow $s_i = 1$). To see this, let $N = \prod p^{a(p)}$ be the canonical decomposition of N into prime powers and define

$$s_k = \prod\{p : a(p) \equiv k \pmod{m}\},$$

for $1 \leqslant k \leqslant m-1$, where $s_k = 1$ if the product is empty. Clearly $N/(s_1 s_2^2 \cdots s_{m-1}^{m-1})$ is an mth power.

For convenience, write P_m for the set of $(m-1)$-tuples $(s_1, \ldots, s_{m-1})$ with $s_1, \ldots, s_{m-1}$ positive, square-free, and pairwise relatively prime. Now suppose that (x, y, z) is a solution of

$$x^2 + y^m = z^{2n} \tag{1}$$

with $xyz \neq 0$, $x > 0$, and $z > 0$. Then one can write

$$z^n + x = s_1 s_2^2 \cdots s_{m-1}^{m-1} u^m \tag{2}$$

with $(s_1, \ldots, s_{m-1})$ in P_m and $u > 0$.

Since (1) can be written as

$$(z^n + x)(z^n - x) = y^m, \tag{3}$$

the decomposition in (2) shows that

$$z^n - x = s_1^{m-1} s_2^{m-2} \cdots s_{m-1} v^m \tag{4}$$

and

$$y = s_1 s_2 \cdots s_{m-1} u\, v, \tag{5}$$

for some integer v with $\operatorname{sgn} v = \operatorname{sgn} y$.

Changing (x, y, z) to $(k^{mn}x, k^{2n}y, k^m z)$ multiplies both $z^n \pm x$ by the factor $k^{mn} = (k^n)^m$. Thus $s_1, \ldots, s_{m-1}$ are determined by the class of the solution. The class of the solution also determines the ratio $u/v = a/b$, say, where a and b are relatively prime and $a > 0$. Since we have assumed x and z to be positive, we have $z^n + x > |z^n - x|$, so (2) and (4) give the condition

$$s_1 s_2^2 \cdots s_{m-1}^{m-1} a^m > s_1^{m-1} s_2^{m-2} \cdots s_{m-1} |b|^m. \tag{6}$$

Conversely, given any $(s_1, \ldots, s_{m-1})$ in P_m and any relatively prime a and b with $a > 0$ which satisfy (6), we will show that (1) has a unique primitive solution satisfying (2), (4), and (5), with $u/v = a/b$. This will substantiate our claim that there are infinitely many solutions and give a method for finding all such solutions.

Thus, write $u = aw$ and $v = bw$ where $w > 0$ is to be determined. Then (2) and (4) give

$$2x = \left(s_1 s_2^2 \cdots s_{m-1}^{m-1} a^m - s_1^{m-1} \cdots s_{m-1} b^m\right) w^m =: A w^m, \tag{7}$$

and

$$2z^n = \left(s_1 s_2^2 \cdots s_{m-1}^{m-1} a^m + s_1^{m-1} \cdots s_{m-1} b^m\right) w^m =: B w^m, \tag{8}$$

say. Here A and B are constants determined by $s_1, \ldots, s_{m-1}$, a, and b. The condition in (6) insures that A and B are positive.

Since A and B have the same parity, if (8) has a solution (z, w) then (1) has a solution (x, y, z) where x is given by (7) and y by (5).

But the solution of (8) is straightforward. If B is even, write

$$B = 2c_1c_2^2 \cdots c_{n-1}^{n-1}d^n,$$

where $(c_1, \dots, c_{n-1})$ is in P_n. Then (8) becomes

$$z^n = c_1c_2^2 \cdots c_{n-1}^{n-1}d^nw^m. \tag{9}$$

In order for the right member of (9) to be an nth power, w must be of the form

$$w = c_1^{i_1} \cdots c_{n-1}^{i_{n-1}}k^n, \tag{10}$$

where

$$mi_j + j \equiv 0 (\text{mod } n), \tag{11}$$

for $j = 1, 2, \dots, n-1$. Since m and n are relatively prime, this has a unique solution i_j with $1 \leqslant i_j \leqslant n-1$ for each j.

The choice $k = 1$ in (10) clearly gives a primitive solution of (9) and a corresponding primitive solution of (1).

If B is odd, write instead

$$B = c_1c_2^2 \cdots c_{n-1}^{n-1}d^n \tag{12}$$

so (8) becomes

$$2z^n = c_1c_2^2 \cdots c_{n-1}^{n-1}d^nw^n \tag{13}$$

for which the general solution has

$$w = 2^{i_0}c_1^{i_1} \cdots c_{n-1}^{i_{n-1}}k^n, \tag{14}$$

where the i_j for $j \geqslant 1$ satisfy (1) and i_0 is determined by $mi_0 \equiv 1$ (mod n) and $1 \leqslant i_0 \leqslant n-1$.

Example 1. Consider first $x^2 + y^3 = z^4$. Then there are four parameters s_1, s_2, a, and b from which we calculate

$$A, B = s_1s_2^2a^3 \pm s_1^2s_2b^3,$$

and then solve (7) and (8). For example, the choice $s_1 = a = b = 1$ and $s_2 = 4$, which satisfies (6), gives $A, B = 16, 20$. Hence (8) becomes $2z^2 = 20w^3$; that is, $z^2 = 2 \cdot 5w^3$. This has primitive solution $w = 2 \cdot 5$, $z = 2^2 \cdot 5^2$. Hence, $x = Aw^3/2 = 2^4 \cdot 3 \cdot 5^3$ and $y = s_1s_2abw^2 = 4 \cdot 10^2$ giving the primitive solution $(2^4 \cdot 3 \cdot 5^3, 2^4 \cdot 5^2, 2^2 \cdot 5^2) = (6000, 400, 100)$ of (1).

Example 2. Consider $x^2 + y^4 = z^6$. Now there are five parameters s_1, s_2, s_3, a, and b. From these one computes

$$A, B = s_1s_2^2s_3^3a^4 \pm s_1^3s_2^2s_3b^4.$$

If, for example, $s_1 = s_2 = s_3 = b = 1$ and $a = 2$, then A, $B = 15$, 17 so (8) becomes $2z^3 = 17w^4$. This has primitive solution $w = 2 \cdot 17^2$, $z = 17^3$ and hence $x = 2^3 \cdot 15 \cdot 17^8$, $y = 2^3 \cdot 17^4$ giving the primitive solution $(2^3 \cdot 15 \cdot 17^8$, $2^3 \cdot 17^4$, $2 \cdot 17^3)$ of (1).

The choice $s_1 = 2, s_2 = 3, s_3 = 5, a = b = 1$ gives $A, B = (2 \cdot 3^2 \cdot 5)(5^2 \pm 2^2)$. So $B = 2 \cdot 3^2 \cdot 5 \cdot 29$ and hence (8) becomes $z^3 = 3^2 \cdot 5 \cdot 29 \cdot w^4$. This has primitive solution $w = 3 \cdot 5^2 \cdot 29^2$, $z = 3^2 \cdot 5^3 \cdot 29^3$. Since $A = 2 \cdot 3^3 \cdot 5 \cdot 7$ we have $x = Aw^4/2 = 3^7 \cdot 5^9 \cdot 7 \cdot 29^8$ and $y = 2 \cdot 3 \cdot 5w^2 = 2 \cdot 3^3 \cdot 5^5 \cdot 29^4$, giving the primitive solution $(3^7 \cdot 5^9 \cdot 7 \cdot 29^8, 2 \cdot 3^3 \cdot 5^5 \cdot 29^4, 3^2 \cdot 5^3 \cdot 29^3)$.

Concluding remarks. Mathieu's solutions to $x^2 + y^3 = z^4$ are suggested by the well known identity

$$1^3 + 2^3 + \cdots + n^3 = \left(n(n+1)/2\right)^2.$$

Thus, if one takes $x = n(n-1)/2$, $y = n$ and $z^2 = n(n+1)/2$, then $x^2 + y^3 = z^4$. In order for z to be an integer, one must solve $2z^2 = n(n+1)$ which is essentially a Pell equation and thus has infinitely many solutions.

On a recent visit to U.B.C., Edwin Hewitt showed me Mathieu's solution and mentioned that a computer search had revealed many other solutions not of this form. An attempt to understand this led to the above method of finding all solutions and its generalization to $x^2 + y^m = z^{2n}$.

A perusal of *Reviews in Number Theory* unearthed the paper of Kiss which happens to be written in Romanian but is nevertheless rather easily read. Although Kiss does not obtain all solutions of $x^2 + y^3 = z^4$, his result that there are infinitely many solutions with x, y and z relatively prime is stronger than the statement that there are infinitely many primitive solutions. Thus it is understandable that more difficult methods are needed.

We do not know whether the more general equation $x^2 + y^m = z^{2n}$ has infinitely many solutions with x, y and z relatively prime.

Acknowledgements. I would like to thank Edwin Hewitt for posing the problem of completely solving $x^2 + y^3 = z^4$. The above research was supported in part by NSERC.

REFERENCES

1. L. E. Dickson, History of the Theory of Numbers, Volume II, Chelsea, 1952.
2. E. Kiss, Rezolvarea in numere naturale a ecutiei diofantiene $x^2 + y^3 = z^4$, Studia Univ. Babes-Bolyai, *Ser. I Math. Phys.* no. 1 (1960), 15–19 (MR 26 #1279).

Modular Arithmetic in the Marketplace

JOSEPH A. GALLIAN AND STEVEN WINTERS
Department of Mathematics and Statistics, University of Minnesota, Duluth, MN 55812

1. Introduction. It is not surprising that there are many schemes that utilize modular arithmetic to append a check digit to product identification numbers for error detection. Two of the better known schemes—the ZIP code bar code and the International Standard Book Number (ISBN)—have been the subjects of articles in the *UMAP Journal* [2], [7]. The schemes used for the Universal Product Code (UPC) and passports are described in [6] and [1].

What is surprising to us is the diversity of the methods in use and the fact that some of them are poorly conceived! In this note we examine several of these schemes.

2. Check digits: from money orders to library books. We begin with the least effective of the methods we have found and work our way up to the best one. The Postal Service's money-order identification number consists of ten digits and a check digit. The check digit is the remainder modulo 9 of the 10-digit number. In contrast to the schemes mentioned in the introduction, this method does not detect all single errors! Indeed, excluding the check digit, a substitution of a 9 for a 0 or vice versa goes undetected. All single-digit errors involving the check digit are detectable. Thus the single-digit error-detection rate for this method is 970/990 or 98.0%. (We assume all errors are equally likely.) Moreover, the only transposition errors detected by this method are those involving the check digit. That is, an error resulting from the transposition of two consecutive digits such as $\ldots 53\ldots$ instead of $\ldots 35\ldots$ is undetected while $\ldots 53$ instead of $\ldots 35$ is detected. Because the digit 9 can never occur as a check digit, this method detects $9 \cdot 10^9 - 8$ of a total of $90 \cdot 10^9 - 8$ possible transposition errors for a rate of 10.0%.

A similar and equally ineffective method is employed on VISA traveler's checks. There the check digit is the additive inverse modulo 9 of the remainder upon division by 9.

Federal Express, airline companies, and the United Parcel Service use a method that is slightly less effective for detecting single errors but fairly effective for detecting transposition errors in their identification numbers. The U.P.S. identification number, for instance, consists of nine digits plus a check digit. The check digit is the remainder modulo 7 of the 9-digit number. Of course, any substitution of b for a in the first nine digits where $|a - b| = 7$ will go undetected. The single-error detection rate for this method is 846/900 or 94.0%. This method detects transposition errors at the rate of $(762 \cdot 10^7 - 5)/(810 \cdot 10^7 - 5)$ or 94.1%.

The Chemical Abstract Service assigns chemicals a registry number together with a check digit calculated in the following way. The number $a_1 a_2 \cdots a_k$ ($k \leqslant 7$) has appended the check digit $(a_1, a_2, \ldots, a_k) \cdot (k, k-1, \ldots, 2, 1) \bmod 10$. All single errors in positions with weighting factors 1, 3, or 7 as well as the check digit position are detected; errors of the form $a \to b$ where $|a - b| = 5$ go undetected in positions

with weighting factors 2, 4, or 6; errors of the form $a \to b$ where $|a - b|$ is even go undetected in the position with a weighting factor of 5. For $k = 7$, this yields a single-error detection rate of 65/72 or 90.3%. On the other hand, *all* transposition errors not involving the check digit are detected. The errors of the type $\dots ab \to \dots ba$ go undetected when $(a_1, a_2, a_3, a_4, a_5, a_6) \cdot (7, 6, 5, 4, 3, 2) \equiv 5 \bmod 10$. Since the probability that this dot product is any particular digit is .1, this yields a transposition error detection rate of 62/63 or 98.4%.

Identification numbers for banks (as appearing on checks, for instance) have eight digits, $a_1 a_2 \cdots a_8$, and a check digit

$$c = (a_1, a_2, \dots, a_8) \cdot (7, 3, 9, 7, 3, 9, 7, 3) \bmod 10.$$

For example, the bank number 09190204 gives $c = 0 + 27 + 9 + 63 + 0 + 18 + 0 + 12 = 129 \equiv 9 \bmod 10$. This method detects 100% of all single errors and 80/90 or 88.9% of all transposition errors. In particular, the only undetectable transpositions are those of the form $ab \to ba$ where $|a - b| = 5$. The advantage this weighting scheme has over one involving just two distinct weighting factors such as $(7, 3, 7, 3, 7, 3, 7, 3)$ is that the former will detect most errors of the form $\dots abc \dots \to \dots cba \dots$ while the latter will detect no errors of this form except those involving the check digit.

The most sophisticated method we have found in use is the "Code-a-bar" system used by many libraries. Here each thirteen digit identification number $a_1 a_2 \cdots a_{13}$ is assigned the check digit

$$-(a_1, a_2, \dots, a_{13}) \cdot (2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2) - r \bmod 10,$$

where r is the number of the digits among $a_1, a_3, a_5, a_7, a_9, a_{11}, a_{13}$ greater than or equal to 5. For example, the identification number 3125600196431 yields the check digit

$$-(6 + 1 + 4 + 5 + 12 + 0 + 0 + 1 + 18 + 6 + 8 + 3 + 2) - 2 = -68 \equiv 2 \bmod 10.$$

This method detects all single errors and all transposition errors except $09 \leftrightarrow 90$. Thus the detection rate for transposition errors is 88/90 or 97.8%. As shown by Gumm [5], it is not possible to improve upon these rates with any system that uses conventional techniques based on addition modulo 10.

3. A foolproof method. The highly effective scheme used by libraries raises the interesting question of whether it is possible to devise a method that will detect all single errors and all transposition errors with a single check digit. Actually, the ISBN method achieves this [7]. But it does so in an artificial way by using the character X to represent the possible check number of 10, which consists of *two* characters. (The method involves modulo 11 arithmetic.)

Recently Gumm [3], [4] has discovered a group theoretical method that uses a single check digit and is 100% effective in detecting single errors and transposition erros. To describe this method we need the dihedral group of order 10, D_5, represented in the form shown in TABLE 1, and the permutation $\sigma = (0)(14)(23)(58697)$. To append a check digit to any string of digits we "weight" the digits with powers of σ and, using TABLE 1, multiply them and take the inverse of the product. For example, consider 1793. The check digit is

$$\left(\sigma^4(1) * \sigma^3(7) * \sigma^2(9) * \sigma(3)\right)^{-1} = (1 * 6 * 5 * 2)^{-1} = 4^{-1} = 1.$$

For 17326, we obtain the check digit

$$\left(\sigma^5(1)*\sigma^4(7)*\sigma^3(3)*\sigma^2(2)*\sigma(6)\right)^{-1} = (4*9*2*2*9)^{-1} = 0^{-1} = 0.$$

To see that this scheme detects all single-digit errors we observe that an error-free number $a_n a_{n-1} \cdots a_1 a_0$ (where a_0 is the check digit) has the property that

$$\sigma^n(a_n)*\sigma^{n-1}(a_{n-1})* \cdots *\sigma(a_1)*a_0 = 0$$

and, therefore, any particular factor in this product is uniquely determined by all of the others. Thus a single-digit error does not result in a product of 0. That all transposition errors are detected can be verified by showing that for all distinct a and b, $\sigma(a)*b \neq \sigma(b)*a$. For then, for all i,

$$\sigma^{i+1}(a)*\sigma^i(b) \neq \sigma^{i+1}(b)*\sigma^i(a)$$

and consequently a transposition will not result in a product of 0.

In addition to being foolproof in detecting single errors and transpositions, Gumm's method will detect approximately 90% of all other types of errors.

TABLE 1. The multiplication table of D_5.

*	0	1	2	3	4	5	6	7	8	9
0	0	1	2	3	4	5	6	7	8	9
1	1	2	3	4	0	6	7	8	9	5
2	2	3	4	0	1	7	8	9	5	6
3	3	4	0	1	2	8	9	5	6	7
4	4	0	1	2	3	9	5	6	7	8
5	5	9	8	7	6	0	4	3	2	1
6	6	5	9	8	7	1	0	4	3	2
7	7	6	5	9	8	2	1	0	4	3
8	8	7	6	5	9	3	2	1	0	4
9	9	8	7	6	5	4	3	2	1	0

For the purpose of comparison, we summarize the calculations given above in TABLE 2 and average the two rates. However, this average does not represent the detection rate for either one of a single-digit error or a transposition error unless these two types of errors are equally likely to occur.

TABLE 2. Summary of error-detection rates.

scheme	single error rate	transposition error rate	average
U.S. Postal Service, VISA	98.0%	10.0%	54.0%
Airlines, U.P.S.	94.0%	94.1%	94.1%
Chemical	90.3%	98.6%	94.5%
Bank	100%	88.9%	94.5%
Library	100%	97.8%	98.9%
Gumm	100%	100%	100%

Acknowledgment. The authors are indebted to David Witte and Douglas Jungreis for their assistance. Joseph Gallian was supported by the NSF (Grant Number DMS 8407498). Steven Winters was supported by a University of Minnesota, Duluth, College of Science and Engineering Undergraduate Honors Research stipend.

REFERENCES

1. Steve Connor, The invisible border guard, *New Scientist* (January 5, 1984) 9–14.
2. Joseph A. Gallian, The ZIP code bar code, *The UMAP Journal*, 7(1986) 191–195.

3. H. Peter Gumm, A new class of check-digit methods for arbitrary number systems, *IEEE Transactions on Information Theory*, 31(1985) 102–105.
4. H. Peter Gumm, Data security through check digits, *Statistical Software Newsletter*, 11(1985) 124–127.
5. H. Peter Gumm, Encoding of numbers to detect typing errors, *International Journal of Applied Engineering Education*, 2 (1986) 61-65.
6. Ian D. Rae, Machine readable codes, *New Zealand Mathematics Magazine*, 21(1984) 109–113.
7. Philip M. Tuchinsky, International Standard Book Numbers, *The UMAP Journal*, 5(1985) 41–54.

On a Property of $x^n e^{-x}$

GABRIEL KLAMBAUER
Department of Mathematics, University of Ottawa, Ottawa, Ontario, Canada, K1N 6N5

Let $f(x) = x^n e^{-x}$, where $x \geqslant 0$ and $n \geqslant 2$. The function f is increasing in the interval $(0, n)$ and decreasing in (n, ∞) and has points of inflection for $x = n \pm \sqrt{n}$. My colleague Ian Iscoe of the University of Ottawa asked me to find a simple proof that $f(n+\sqrt{n}) > f(n-\sqrt{n})$. I wish to show here that, more generally, $f(n+h) > f(n-h)$, where $0 < h < n$.

Indeed, the inequalities

$$\frac{(n+h)^n e^{-n-h}}{(n-h)^n e^{-n+h}} > 1 \quad \text{and} \quad \ln\frac{n+h}{n-h} > \frac{2h}{n}$$

are equivalent. The latter inequality can be proved as follows. Note that

$$\ln\{(n+h)/(n-h)\}$$

is the area of the region H below the hyperbola $y = 1/x$ and above the interval $[n-h, n+h]$ on the x-axis. On the other hand, $2h/n$ can be interpreted as the area of the trapezoidal region T below the tangent line to $y = 1/x$ at $x = n$, the midpoint of the interval $[n-h, n+h]$, and above the interval $[n-h, n+h]$ on the x-axis. Observing that $y = 1/x$ is concave upward for $x > 0$, it is apparent that the region T is contained in the region H and so the area of T is smaller than the area of H.

Noncentral Difference Quotients and the Derivative

P. P. B. EGGERMONT
Department of Mathematical Sciences, University of Delaware, Newark, DE 19716

We all have proved at one time or another the result that if a function f, defined on a neighborhood of the origin, is differentiable at the origin, then the central difference quotients converge, i.e.,

$$\lim_{h\to 0}\frac{f(h)-f(-h)}{2h}$$

exists and equals $f'(0)$, but that the converse is not true in general, the counterexam-

ple being $f(x) = |x|$ (see, e.g., [**2**, Problem 27.0] or [**1**, Problem 5.20]). What can be said about noncentral difference quotients? For instance, does the existence of the limit

$$\lim_{h \to 0} \frac{f(h) - f(-2h)}{3h} \tag{1}$$

imply that f is differentiable at the origin? To avoid trivialities, we must assume that f is continuous at the origin. Note that $f(x) = |x|$ is *not* a counterexample in the noncentral case.

The surprise is that the existence of the limit (1) does indeed guarantee the differentiability of f. This simple result does not appear to be well known, which is amazing since the proof is entirely within the scope of standard advanced calculus. Results in the literature mostly deal with consequences of one-sided limits, (see, e.g., [**3**], [**4**], [**5**]). As a general reference, see [**6**]. We prove the following.

THEOREM. *Let f, defined on a neighborhood of* 0, *be continuous at* 0, *and let $a \neq \pm 1$. If the limit*

$$\lim_{h \to 0} \frac{f(h) - f(ah)}{(1-a)h} = L$$

exists, then f is differentiable at the origin.

Proof. Let

$$D(a, h) = \frac{f(h) - f(ah)}{(1-a)h}.$$

The proof is simplified if we could take $L = 0$ and $0 < a < 1$. Can we? Observe that (i) if $L \neq 0$, we can consider $f(x) - Lx$; (ii) if $a = 0$, we just have the *definition* of differentiability; (iii) if $|a| > 1$, we can take as our starting point

$$\lim_{h \to 0} D(a, h/a) = \lim_{h \to 0} \frac{f(h) - f(h/a)}{(1 - a^{-1})h},$$

for which the new constant a is the old a^{-1}; (iv) if $a < 0$, we can consider $D(a^2, h)$ since

$$D(a^2, h) = \frac{D(a, h)}{1+a} + \frac{aD(a, ah)}{1+a}.$$

We may thus assume that $L = 0$ and $0 < a < 1$.

Observe that

$$\sum_{j=0}^{n-1} a^j D(a, a^j h) = \frac{f(h) - f(a^n h)}{(1-a)h}. \tag{2}$$

Let $\varepsilon > 0$. There exists a $\delta > 0$ such that for all h with $0 < |h| < \delta$ we have $|D(a, h)| < \varepsilon$. But then for all $j \geqslant 0$ we have $|D(a, a^j h)| < \varepsilon$ as well. Consequently, for all $0 < |h| < \delta$ and for all $n \geqslant 1$, using (2),

$$\left| \frac{f(h) - f(a^n h)}{(1-a)h} \right| < (1 - a^n)\varepsilon/(1-a),$$

from which we obtain by the triangle inequality that

$$\left|\frac{f(h)-f(0)}{h}\right| < \varepsilon + \left|\frac{f(a^n h)-f(0)}{h}\right|.$$

Since f is continuous at 0, and $0 < a < 1$, we can choose n so large that $|f(a^n h) - f(0)| < \varepsilon|h|$. Note that n may depend on h. Thus, we have shown that for all h with $0 < |h| < \delta$,

$$\left|\frac{f(h)-f(0)}{h}\right| < 2\varepsilon$$

and we are done. □

In the same vein, we may ask under which conditions on the function $g(h)$ is the differentiability of f at 0 implied by the existence of the limit

$$\lim_{h \to 0} \frac{f(h) - f(g(h))}{h - g(h)}.$$

A sufficient condition appears to be that $|g(h)| < |h|$ and $g(h)g(g(h))$ always has the same sign. (Just imitate the proof of the theorem with $a^n h$ replaced by $g_n(h)$, the n-fold composition of g with itself.) So the anwer to the above question is yes for $g(h) = \sin(h)$, but what if $g(h) = s(h)\sin h$ where $s(h)$ takes on the values $+1$ and -1 in some "random" manner?

Acknowledgements. Thanks are due to the referee for a job well done.

REFERENCES

1. T. M. Apostol, Mathematical Analysis, Addison-Wesley, Reading, Mass., 1957.
2. R. G. Bartle, The Elements of Real Analysis, second ed., John Wiley and Sons, New York, 1964.
3. G. H. Sindalovskii, On a generalization of derived numbers (Russian), *Izv. Akad. Nauk SSSR Ser. Mat. 24*, 707–720 (1960) MR 23-A1760.
4. V. Starcev, A certain generalization of the concept of symmetric continuity and symmetric differentiability (Russian), *Izv. Vyss. Ucebn. Zaved. Matematika*, 1971, no. 3 (106) 92–100 MR 45 #2112.
5. S. Valenti, Sur la derivation k-pseudo-symmetrique des fonctions numeriques, *Fund. Math.*, 74 (1972) 147–152.
6. P. Ver Eecke, Fondaments du Calcul Differentiel, Presses Universitaires de France, Paris, 1983.

PROBLEMS AND SOLUTIONS

For instructions about submitting ***proposed*** *problems for publication in this department see the inside front cover. Please include solutions, relevant references, etc.*

As asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

Solutions *should be sent to the address given on the inside front cover. Two copies suffice.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

For instructions about submitting solutions of Problems, which should be mailed before October 31, 1988, see the inside front cover. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label if an acknowledgment is desired.

E 3269. *Proposed by M. J. Pelling, University College, London, England.*

For what positive integers n does there exist a permutation $(x_1, x_2, \dots, x_n)$ of $(1, 2, \dots, n)$ such that the differences $|x_k - k|$, $1 \leqslant k \leqslant n$, are all distinct?

E 3270. *Proposed by Berndt Lindström, University of Stockholm, Sweden.*

Determine those positive rational numbers m such that

$$\frac{1}{\pi}\arctan\sqrt{m}$$

is also rational.

E 3271. *Proposed by Walter Noll and Joseph C. Keane, Carnegie Mellon University.*

For each $r \in \mathbb{Q}$ (the set of all rational numbers), denote the denominator of the representation of r as a reduced fraction by $\delta(r)$. For each $\alpha \in \mathbb{R}$ (the set of all real numbers) consider

$$\sigma(\alpha) := \sum_{r \in S} (\delta(r))^{-\alpha},$$

where

$$S := \{r \in \mathbb{Q} \mid 0 < r \leqq 1\}.$$

Show that $\sigma(\alpha)$ is finite if and only if $\alpha > 2$ and evaluate $\sigma(\alpha)$ when $\alpha > 2$.

E 3272. *Proposed by K. I. Appel and C. G. Jockusch, Jr., University of Illinois at Urbana-Champaign.*

It is well known that $\frac{1}{3}(n^3 + 2n - 3)$ multiplications/divisions are required to find the determinant of an $n \times n$ matrix by Gaussian elimination, while obviously the permutation expansion requires $(n-1)n!$ multiplications. Obtain a formula in closed form for the number a_n of multiplications to evaluate an $n \times n$ determinant by the cofactor method. The formula should not involve summations but may use the greatest integer function.

E 3273. *Proposed by Orrin Frink, Pennsylvania State University.*

Give appropriate conditions under which a simple closed curve in the plane contains three points which form the vertices of an equilateral triangle. For the corresponding problem for the square see Problem 4325 [1949, 39; 1950, 423] and Victor Klee, Some Unsolved Problems in Plane Geometry, *Math. Magazine*, 52 (1979) 131–145.

E 3274. *Proposed by David Newman, Beer Sheva, Israel.*

Fix a positive integer k. Define $f(n)$ on positive integers by $f(n) = 1$ for $n \leqslant k+1$ and $f(n) = f(f(n-1)) + f(n - f(n-1))$ for $n > k+1$. Define the sequence F_m by $F_m = 1$ for $m \leqslant k$ and $F_m = F_{m-1} + F_{m-k}$ for $m > k$. (Note that the F_m are powers of 2 when $k = 1$ and ordinary Fibonacci numbers when $k = 2$.)

(a) Prove that $f(n) - f(n-1)$ is 0 or 1 for all n, and that $f(n)$ is unbounded.

(b) Prove that $f(F_{m+k}) = F_m$ for $m \geqslant 1$.

E 3275. *Proposed by Thomas J. Laffey, University College, Dublin and Desmond MacHale, University College, Cork, Ireland.*

(a) Suppose R is an associative ring, not necessarily with unity, in which every element is a square. If squaring is a ring homomorphism, prove that R is commutative.

(b) Suppose R is an associative ring and suppose there is a function f from R onto R which satisfies $f(x+y) = f(x) + f(y)$ and $f(xy) = (xy)^2$ for all $x, y \in R$. Must R be commutative?

SOLUTIONS OF ELEMENTARY PROBLEMS

A Generalized Fibonacci Recurrence

E 3130 [1986, 131]. *Proposed by Mark Kantrowitz (student), Brookline, MA.*

Let $\{H_n\}$ be a generalized Fibonacci sequence, i.e. H_1 and H_2 are arbitrary integers and, for $n > 2$, $H_n = H_{n-1} + H_{n-2}$.

(a) Find T, in terms of H_1 and H_2, such that $H_{2n}H_{2n+2} + T$, $H_{2n}H_{2n+4} + T$, $H_{2n-1}H_{2n+1} - T$, and $H_{2n-1}H_{2n+3} - T$ are all perfect squares.

(b) Prove that T is unique.

Solution by Samuel W. Bent, Dartmouth College, Hanover, NH.

(a) Applying the Fibonacci recurrence and induction, we have

$$\begin{bmatrix} H_{n+2} & H_{n+1} \\ H_{n+1} & H_n \end{bmatrix} = \begin{bmatrix} H_{n+1} & H_n \\ H_n & H_{n-1} \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} H_3 & H_2 \\ H_2 & H_1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^{n-1}.$$

Letting $T = H_3H_1 - H_2^2$ and taking determinants converts this to $H_{n+2}H_n + (-1)^nT = H_{n+1}^2$, which yields the first and third results for even and odd values of n, respectively.

Similarly, since $H_{n+2} - H_n = H_{n+1}$ and $2H_{n+2} - H_n = H_{n+2} + H_{n+1} = H_{n+3}$, we have

$$\begin{bmatrix} H_{n+4} & H_{n+2} \\ H_{n+2} & H_n \end{bmatrix} = \begin{bmatrix} H_{n+3} & H_{n+1} \\ H_{n+1} & H_{n-1} \end{bmatrix} \cdot \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} H_5 & H_3 \\ H_3 & H_1 \end{bmatrix} \cdot \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}^{n-1}.$$

Using the further identity

$$\begin{bmatrix} H_5 & H_3 \\ H_3 & H_1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} H_3 & H_2 \\ H_2 & H_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

and taking determinants, we have $H_nH_{n+4} + (-1)^nT = H_{n+2}^2$, which yields the second and fourth results for even and odd values of n, respectively.

(b) It is easy to verify by induction that $H_{n+1} = F_{n-1}H_1 + F_nH_2$, where $\{F_n\}$ are the usual Fibonacci numbers. The well-known identity

$$F_n = \frac{1}{\sqrt{5}}(\phi^n - \hat{\phi}^n),$$

where $\phi = (1 + \sqrt{5})/2$ and $\hat{\phi} = (1 - \sqrt{5})/2$, implies

$$H_{n+1} = \frac{1}{\sqrt{5}}(H_1/\phi + H_2)\phi^n + O(|\hat{\phi}|^n).$$

Furthermore, $(H_1/\phi + H_2) \neq 0$, since ϕ is irrational.

Suppose that T' is a solution to the problem, so that $H_{n+2}H_n + (-1)^nT' = b_{n+1}^2$ for some sequence of integers $\{b_n\}$. Then $H_{n+1}^2 - b_{n+1}^2 = (-1)^n(T - T')$. The left side is either zero or has magnitude at least $2|H_{n+1}| - 1$ (the distance from H_{n+1}^2 to the nearest square). But $2|H_{n+1}| - 1 = c\phi^n + O(1)$ for some nonzero constant c, and for large enough n this exceeds the constant $|T - T'|$. Thus for large n the left side must be zero, which implies $T - T' = 0$.

Editorial Comment. The problem has appeared before, as noted by several readers. G. Philippou cited Problem B-549 in *The Fibonacci Quarterly* [1985, 182; 1986, 184] and H. V. Krishna cited a paper of A. F. Horadam, "A generalized Fibonacci sequence," *Amer. Math. Monthly*, 68 (1961) 455–459.

Solved also by 18 other readers and the proposer. Several incorrect or partial solutions were received.

Trapezoids in a Polygon

E 3157 [1986, 482]. *Proposed by Liviu I. Nicoaescu, Iassy, Romania.*

How many sets of four distinct points forming the vertices of a trapezoid are there if the points are chosen from the vertices of a regular n-gon ($n \geqslant 4$)?

Solution by Dennis Estes and Charles Lanski, University of Southern California.

The answer is $n\binom{(n-1)/2}{2}$ if n is odd and $(n-3)\binom{n/2}{2}$ if n is even. Any trapezoid determined by four vertices of a regular n-gon is defined by two parallel chords joining vertices, so we count the number of such pairs of parallel chords. Any chord joining two vertices is parallel to a side or is perpendicular to a diameter (if n is even); we organize the counting using this fact.

If n is odd, then each side has $(n-1)/2$ chords parallel to it (including itself), giving $\binom{(n-1)/2}{2}$ pairs of chords. The sets of chords corresponding to distinct sides are disjoint, so there are $n\binom{(n-1)/2}{2}$ pairs of parallel chords. We claim each pair generates a distinct trapezoid. Two pairs generate the same trapezoid if and only if they form a parallelogram. This requires the chords in a pair to have the same length, meaning their endpoints differ by the same number of vertices around the n-gon. Hence, vertices of an n-gon can determine a parallelogram only if n is even.

If n is even, then there are $\binom{n/2}{2}$ pairs of chords parallel to each side. Opposite sides yield the same pairs, so there are $(n/2)\binom{n/2}{2}$ pairs of chords parallel to sides. Each diagonal has $(n-2)/2$ chords perpendicular to it, so altogether there are $(n/2)\binom{n/2-1}{2}$ pairs of chords perpendicular to diameters. As before, two pairs generate the same trapezoid if and only if they form a parallelogram. In terms of vertex distance around the n-gon, the lengths of these sides must be i, j, i, j in order, with $i+j=n/2$. There are $n/2-1$ such parallelograms starting at any vertex. The n vertices give rise to $(n/2-1)n$ parallelograms, each counted four times, once for each of its four vertices. Thus there are

$$(n/2-1)n/4=\binom{n/2}{2}$$

distinct parallelograms. Eliminating the duplication, we have

$$(n/2)\binom{n/2}{2}+(n/2)\binom{n/2-1}{2}-\binom{n/2}{2}=(n-3)\binom{n/2}{2}$$

trapezoids when n is even.

Also solved by J. Cohen, R. B. Eggleton (Australia), N. Felsinger, W. Janous (Austria), J. M. Rojas (student), K. L. Stellmacher, J. T. Ward, H.-F. Yeung (Australia), and the proposer. Five incorrect solutions were received.

A Natural Limit

E 3159 [1986, 565]. *Proposed by Emeric Deutsch, Polytechnic Institute of New York.*

Define a sequence (p_n) by $p_0=1$ and

$$p_n+\frac{1}{1!}p_{n-1}+\frac{1}{2!}p_{n-2}+\cdots+\frac{1}{(n-1)!}p_1+\frac{1}{n!}p_0=1.$$

Show that (p_n) is convergent and find its limit.

Solution I by Jonathan M. Borwein, Dalhousie University, Halifax, Nova Scotia.

Let $p(x)=\Sigma p_n x^n$ be the ordinary generating function for (p_n). The recurrence given for p_n is the nth term of the series product $p(x)e^x=\Sigma x^n=1/(1-x)$. Hence, $p(x)=e^{-x}/(1-x)$ and $p_n=\Sigma_{k=0}^{n}(-1)^k/k!$, which has limit $1/e$.

Solution II by Paul Cull, Oregon State University, Corvallis, OR.

Let $g_n(x) = \sum_{k=0}^{n} p_{n-k} x^k/k!$. The recurrence implies $g_n(1) = 1$ for all n, and the definition of g_n implies $g'_{n+1}(x) = g_n(x)$. The solution to this recursive differential equation is $g_n(x) = \sum_{k=0}^{n}(x-1)^k/k!$. Setting $x = 0$, we have $p_n = \sum_{k=0}^{n}(-1)^k/k!$, which has limit $1/e$.

Solution III by G. Behrendt, Universität, Tübingen, West Germany.

Let D_n be the number of derangements of n elements, i.e. the number of permutations of $1, \dots, n$ leaving no element in its original position. There are $\binom{n}{k} D_{n-k}$ permutations leaving exactly k elements fixed. Since every permutation leaves some number of elements fixed, we have $\sum_{k=0}^{n}\binom{n}{k} D_{n-k} = n!$. Dividing by $n!$ shows that the ratios $D_n/n!$ satisfy the recurrence for the p_n; also $D_0 = k_0 = 1$. By the inclusion-exclusion principle we know that $D_n/n! = \sum_{k=0}^{n}(-1)^k/k!$, which has limit $1/e$.

Editorial comment. S. H. E. Hou noted that another solution appears in D. E. Knuth, *The Art of Computer Programming*, Vol. 1, Addison-Wesley (1973), 62–63.

Also solved by over 75 other readers and the proposer.

A Varied Menu

E 3169 [1986, 650]. *Proposed by Dean S. Clark, University of Rhode Island.*

A gentleman, who decides randomly what to eat for dinner so long as it isn't the same thing two nights in a row, will return to his favorite restaurant next Friday. Is it more likely that he will order (a) Chicken Mediterranean or (b) the dinner he had (which he cannot remember) when he first visited the restaurant? How much more likely?

Solution by David Callan, University of Bridgeport.

It depends on whether he has visited the restaurant an even or odd number of times. The choice of dishes on successive visits forms a finite Markov chain, since each choice depends only on the previous one. Assuming there are n dishes available, the n by n transition probability matrix P has zeroes on the diagonal and $1/(n-1)$ everywhere else. So $P = (n-1)^{-1}(J - I)$, where J is the n by n matrix of ones. Replacing $(n-1)^{-1}$ by $1/n + 1/[n(n-1)]$ yields $P = n^{-1}J - (n-1)^{-1}(I - n^{-1}J)$, which is the basis step for an elementary inductive proof that

$$P^r = \frac{1}{n}J + \frac{(-1)^r}{(n-1)^r}(I - n^{-1}J).$$

The induction step uses the fact that $J^2 = nJ$.

From this, the probability of repeating a dish r visits later is $1/n + (-1)^{r-1}/[n(n-1)^{r-1}]$. Since he cannot remember the dish he had on his first visit, the probability of (a) is $1/n$. The probability of (b) differs from this by $1/[n(n-1)^{r-1}]$, and is larger on even visits and smaller on odd visits.

Editorial comment. Several solvers interpreted the statement of the problem to exclude the possibility that the dinner ordered on the first visit was Chicken

Mediterranean. This leads to the same qualitative result as above, but the difference in probabilities becomes $1/(n-1)^r$.

Also solved by S. F. Barger, J. W. Grossman, R. Hill (England), C. Hurd, O. P. Lossers (The Netherlands), J. T. Ward, and the proposer.

ADVANCED PROBLEMS

For instructions about submitting solutions of Problems, which should be mailed before October 31, 1988, see the inside front cover. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label if an acknowledgment is desired.

6574. *Proposed by Moshe Laub, Jerusalem, Israel.*

Put $M_r = \sum_{j=1}^r j^2$ for $r = 1, 2, \ldots$. If $n \in N$ (the set of positive integers), define $m(n)$ to be the unique positive integer r such that $M_r \leqslant n < M_{r+1}$.

(a) Let S_1 be the set of $n \in N$ such that n is expressible as a sum of $m(n)$ distinct squares of positive integers. Prove that S_1 has asymptotic density one but that $N \setminus S_1$ is infinite.

(b) Let S_2 be the set of $n \in N$ such that n^2 is expressible as a sum of $m(n^2)$ distinct squares of positive integers. Prove that S_2 has asymptotic density one. *Is $N \setminus S_2$ infinite?

6575. *Proposed by Richard Stanley, Massachusetts Institute of Technology.*

Let $x_1, \ldots, x_n, y_1, \ldots, y_n$ be variables and for $1 \leqslant i \leqslant j \leqslant n$ set $X_{ij} = x_i + x_{i+1} + \cdots + x_j$ and $Y_{ij} = y_i + y_{i+1} + \cdots + y_j$. Also set $X_{i+1,i} = Y_{i+1,i} = 0$, and any product $\prod_h^{h-1}$ equal to 1. Prove the identity

$$\sum_{i=1}^{n} \frac{\prod_{j=1}^{i}(X_{ji} + Y_{j,i-1}) \cdot \prod_{k=1}^{n+1-i}(X_{i+1,n+1-k} + Y_{i,n+1-k})}{\prod_{r=1}^{i-1}(X_{r+1,i} + Y_{r,i-1}) \cdot \prod_{s=1}^{n-i}(X_{i+1,n+1-s} + Y_{i,n-s})} = \sum_{i \leqslant j} x_i y_j.$$

For instance, when $n = 2$ we get

$$\frac{x_1(x_2 + y_1 + y_2)y_1}{x_2 + y_1} + \frac{(x_1 + x_2 + y_1)x_2 y_2}{x_2 + y_1} = x_1 y_1 + x_1 y_2 + x_2 y_2.$$

SOLUTIONS OF ADVANCED PROBLEMS

Maximizing an Integral over a Diagonal Strip

6521 [1986, 485]. *Proposed by C. S. Gardner and C. Radin, University of Texas at Austin.*

Let a, b be fixed, with $0 < a < b$, and let f be a variable function, subject to

$$f(x) \geqslant 0, \; -\infty < x < \infty, \text{ and } \int_{-\infty}^{\infty} f(x)\,dx = 1.$$

What is the maximum value of

$$V(f) = \int_{-\infty}^{\infty} \int_{y+a}^{y+b} f(x) f(y)\,dx\,dy?$$

Solution by Robert B. Israel, University of British Columbia, Vancouver. The maximum value of $V(f)$ is $n/(2n+2)$ where n is the greatest integer strictly less than b/a. We first show that this value can be attained. Choose c and $\delta > 0$ such that $a < c - 2\delta$ and $nc + 2\delta < b$. Then for any $x_0, \ldots, x_n$ with $|x_k - kc| < \delta$, we will have $a < |x_i - x_j| < b$ whenever $i \neq j$. Let f be a function supported in the union of the intervals $(kc - \delta, kc + \delta)$ for $k = 0$ to n, with

$$\int_{kc-\delta}^{kc+\delta} f(x)\,dx = \frac{1}{n+1}.$$

Then

$$\begin{aligned} V(f) &= \sum_{i=0}^{n} \sum_{j=i+1}^{n} \int_{ic-\delta}^{ic+\delta} \int_{jc-\delta}^{jc+\delta} f(x) f(y)\,dx\,dy \\ &= \frac{1}{2}\left(1 - \sum_{i=0}^{n}\left(\int_{ic-\delta}^{ic+\delta} f(x)\,dx\right)^2\right) = \frac{n}{2(n+1)}. \end{aligned}$$

Next, we show that this value is an upper bound for $V(f)$. It is easy to see that it suffices to prove this for functions f with bounded support, since, for any f satisfying the requirements, a suitable cutoff and rescaling produces a function g with bounded support and $V(g)$ arbitrarily close to $V(f)$.

Consider two intervals I and J, each of length a or less, and separated by a distance of at least b. We claim that if $\alpha = \int_I f(x)\,dx$ and $\beta = \int_J f(x)\,dx$ are both nonzero, then there is a function g that is equal to f outside I and J, zero on one of the intervals and a positive multiple of f on the other, so that

$$\int_{-\infty}^{\infty} g(x)\,dx = 1 \text{ and } V(g) \geqslant V(f).$$

Let two candidates for g be

$$g_1(x) = \begin{cases} 0 & x \in I, \\ \dfrac{\alpha+\beta}{\beta} f(x) & x \in J, \\ f(x) & \text{otherwise,} \end{cases}$$

and

$$g_2(x) = \begin{cases} \dfrac{\alpha+\beta}{\alpha} f(x) & x \in I, \\ 0 & x \in J, \\ f(x) & \text{otherwise.} \end{cases}$$

Then

$$f=\frac{\beta}{\alpha+\beta}g_1+\frac{\alpha}{\alpha+\beta}g_2.$$

Note that if $y+a<x<y+b$ then x and y are not both in $I\cup J$. If $x\notin I\cup J$ then

$$\begin{aligned}f(x)f(y)&=f(x)\left(\frac{\beta}{\alpha+\beta}g_1(y)+\frac{\alpha}{\alpha+\beta}g_2(y)\right)\\&=\frac{\beta}{\alpha+\beta}g_1(x)g_1(y)+\frac{\alpha}{\alpha+\beta}g_2(x)g_2(y).\end{aligned}$$

A similar argument shows that the extreme left- and right-hand sides of the above equality are equal when $y\notin I\cup J$. Therefore,

$$V(f)=\frac{\beta}{\alpha+\beta}V(g_1)+\frac{\alpha}{\alpha+\beta}V(g_2).$$

Thus at least one of $V(g_1)$ and $V(g_2)$ is greater than or equal to $V(f)$, which proves the claim.

Now let f be any function of bounded support. A finite number of applications of the result of the last paragraph produces a function g with support of diameter at most b, and $V(g)\geqslant V(f)$. Since $(n+1)\ a\geqslant b$, the support of g is contained in the union of $n+1$ intervals $I_0,\dots,I_n$ (from left to right) of length a. Let $p_i=\int_{I_i}g(x)\,dx$. Then

$$V(g)\leqslant\sum_{i=0}^{n}\sum_{j=i+1}^{n}p_ip_j=\frac{1}{2}\left(1-\sum_{i=0}^{n}p_i^2\right)\leqslant\frac{n}{2(n+1)},$$

since the minimum of $\Sigma_{i=0}^{n}p_i^2$ subject to $\Sigma_{i=0}^{n}p_i=1$ is attained with all $p_i=1/(n+1)$.

Editorial Comment. Most akin to Israel's solution was that of O. P. Lossers. The proposers and (independently) Kenneth Schilling used the language of graph theory in essentially the same way to solve the problem. The relevant lemma (in Schilling's notation) is the following. Let G be an undirected graph with e edges, v vertices, and no subgraph that is a complete graph on $n+1$ vertices. Then

$$e\leqslant\frac{n-1}{2n}v^2.$$

This is a weak form of a theorem of Turán, which asserts that

$$e\leqslant\frac{n-1}{2n}v^2-\frac{n}{2}\theta(1-\theta),$$

where $\theta=v/n-\lfloor v/n\rfloor$.

Quotients Which Are Roots of Unity

6523 [1986, 485]. *Proposed by David G. Cantor, University of California, Los Angeles.*

Suppose $f(z)$ is an irreducible polynomial of degree d over the field of rational numbers, and suppose that $f(z)$ has two roots α, β with α/β a primitive nth root of unity. Show that $\phi(n) \leqslant d$.

Solution by I. M. Isaacs, University of Wisconsin, Madison, WI. We are given $f(\alpha) = 0 = f(\beta)$ where f is irreducible over Q and $\alpha/\beta = \varepsilon$, a primitive nth root of unity. Let E be a splitting field for f containing α and β and let $G = \mathrm{Gal}(E/Q)$, the Galois group. Let $H = \mathrm{Gal}(E/Q(\alpha))$ and $K = \mathrm{Gal}(E/Q(\beta))$ and note that the subgroups H and K are conjugate in G since f is irreducible. Also

$$|G : H| = d = |G : K|,$$

where $d = \deg(f)$.

Now let $N = \mathrm{Gal}(E/Q(\varepsilon))$. Since ε is a root of unity, $Q(\varepsilon)$ is Galois over Q and so N is normal in G. Also, G/N is isomorphic to $\mathrm{Gal}(Q(\varepsilon)/Q)$ and so it is abelian of order $\phi(n)$. Now HN and KN are conjugate in G and since G/N is abelian, we conclude that $HN = KN \supseteq K$. We also have $H \cap K = H \cap N$, since an element of H fixes $\alpha = \beta\varepsilon$ and so it fixes β if and only if it fixes ε.

We have

$$|HN : H| \cdot |HN : K| \geqslant |HN : H \cap K| = |HN : H \cap N| = |HN : H| \cdot |H : H \cap N|,$$

and so $|HN : K| \geqslant |H : H \cap N| = |HN : N|$. We conclude that $|K| \leqslant |N|$ and so $d \geqslant \phi(n)$, as required.

Editorial Comment. Modifying the above argument by replacing group indices by the degrees of the corresponding field extensions leads to the following more general result, as was pointed out by Cantor. Let α and β be conjugate algebraic elements over a field k, and suppose K is an abelian extension of k contained in $k(\alpha, \beta)$; then $[K : k] \leqslant [k(\alpha) : k]$. A proof along these lines and a similar generalization was presented by Ze-Li Dou (student, Queens College, CUNY).

Many readers took questionable shortcuts here, and a number of proposed solutions were judged as either incorrect or incomplete.

Also solved by Ze-Li Dou, Edward H. Grossman, David R. Richman, John Henry Steelman, and the proposer.

An Alternating Sum of Products of Beta Random Variables

6524 [1986, 573]. *Proposed by Gérard Letac and Guy Yétèrian, Université Paul Sabatier, Toulouse, France.*

Let p and q be positive numbers, and $\{X_n\}_{n=0}^{\infty}$ a sequence of independent random variables with the same distribution

$$\beta_{p,q}(dx) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1}(1-x)^{q-1}\, dx$$

on $[0,1]$. Find the distribution of

$$\sum_{n=0}^{\infty} (-1)^n X_0 X_1 \cdots X_n.$$

Solution by Western Maryland College Problems Group, Western Maryland College, Westminster. Let

$$Z_m = \sum_{n=0}^{m} (-1)^n X_0 X_1 \cdots X_n,$$

and set $Z = Z_\infty$. We will show that Z has a β density

$$g(t) = \frac{\Gamma(2p+q)}{\Gamma(p)\Gamma(p+q)} t^{p-1}(1-t)^{p+q-1} \text{ on } [0,1].$$

Observe that, since $P\{X_i \leqslant \frac{1}{2}\}$ is constant, the event $\{X_i \leqslant \frac{1}{2}\}$ occurs infinitely often (with probability one). Hence, the $(n+1)$-fold product of X_i contains a factor of an arbitrarily large power of $\frac{1}{2}$ if n is large enough; hence it converges to zero.

Now

$$P[X_n = 1 \text{ for some } n] \leqslant \sum_{n=0}^{\infty} P[X_n = 1] = 0.$$

Thus Z_m converges with probability one because the series is alternating and it fails to decrease with probability zero. It follows that Z_m converges in distribution to Z. Denote by G_m the distribution of Z_m and G the distribution of Z. We can factor out X_0 and write

$$Z_m = X_0\left[1 - \sum_{n=1}^{m} (-1)^{n-1} X_1 \cdots X_n\right] = X_0[1 - Y_{m-1}],$$

where Y_{m-1} is independent of X_0 and has the same distribution as Z_{m-1}. This will give us an integral equation satisfied by G. We write f for the density of X_0.

Now $G_m(t) = P[Z_m \leqq t]$ satisfies

$$\begin{aligned} G_m(t) &= P[X_0(1 - Y_{m-1}) \leqq t] \\ &= 1 - P[Y_{m-1} \leqq 1 - t/X_0] \\ &= 1 - \int_t^1 G_{m-1}(1 - t/s) f(s)\, ds. \end{aligned}$$

By letting m go to infinity we get

$$G(t) = 1 - \int_t^1 G(1 - t/s) f(s)\, ds.$$

This equation has a unique solution for $G \in L^1[0,1]$. To see this consider the operator ϕ defined by

$$\phi[H](t) = 1 - \int_t^1 H(1 - t/s) f(s)\, ds.$$

By a change in the order of integration, and the substitution $u = 1 - t/s$, we find that

$$\begin{aligned} \|\phi[H] - \phi[K]\| &= \left|\int_0^1 \int_t^1 (H(1 - t/s) - K(1 - t/s)) f(s)\, ds\, dt\right| \\ &\leqslant \int_0^1 \int_0^s |H(1 - t/s) - K(1 - t/s)| f(s)\, dt\, ds \\ &= \int_0^1 \int_0^1 |H(u) - K(u)| s f(s)\, du\, ds \\ &= \|H - K\| \mu, \end{aligned}$$

where $\mu = E[X_0] = p/(p+q) < 1$. By the contraction mapping theorem ϕ has a unique fixed point. Also, the equation for G shows that G has a continuous derivative on $(0,1)$. Upon differentiating the equation for G we get an equation satisfied by the density of Z, namely,

$$g(t) = \int_t^1 g(1 - t/s)s^{-1}f(s)\,ds.$$

This must also have a unique solution for g a probability density, because we can integrate to recover the equation for G and then differentiate to get g. It is now a relatively straightforward exercise in integral calculus to show (after some changes of variable) that the $g(t)$ mentioned at the start, i.e., the $\beta(p, p+q)$ density, satisfies this equation.

Editorial Comment. Specialists in probability theory might prefer a short proof given by Norman L. Johnson (University of North Carolina) who quickly obtains the result by taking rth moments of both sides of

$$Z_\infty = X_0(1 - Y_\infty), \tag{$*$}$$

and by using the fact that the product of two mutually independent beta variables with parameters (p, q) and $(p+q, p)$ has a beta distribution with parameters $(p, p+q)$. O. P. Lossers (The Netherlands), whose solution is somewhat similar to the one from Western Maryland, also relies on moments to establish uniqueness. Lossers adds that the product relationship between (p, q), $(p+q, p)$ and $(p, p+q)$ is related to equalities for the order statistics $Y_{k;n}$ from the exponential distribution by

$$\exp(-Y_{k;n}) = X_{n-k,k}.$$

Here $X_{p,q}$ denotes a random variable with distribution $\beta_{p,q}$ and equality means equality in distribution.

The proposers note that for $p = q = 1$, the result is essentially due to T. Ugaheri, On a limit distribution, *Ann. Inst. Statist. Math. Tokyo*, 1 (1950) 157–160.

Also solved by Ignacy Icchak Kotlarski, Kenneth Schilling, Douglas P. Wiens, and the proposers.

A Matrix Whose Cube Is the Identity

6527 [1986, 659]. *Proposed by Nicholas Strauss, Boston University.*

Let $J(m)$ be the $m \times m$ matrix whose (i, j)th entry is

$$\binom{i+j}{j}, \quad 0 \leqslant i, j \leqslant m-1.$$

Show that for all primes p and positive integers n, the matrix $J(p^n)$ is a cube root of unity modulo p.

Solution by Ira Gessel, Brandeis University, Waltham, MA. Let $K(m)$ be the $m \times m$ matrix whose (i, j)th entry is

$$(-1)^j\binom{m-i-1}{j}, \quad 0 \leqslant i, j \leqslant m-1.$$

Let Θ be the linear transformation on polynomials in x of degree less than m

defined by

$$\Theta(f(x)) = (1 - x)^{m-1} f\left(\frac{1}{1 - x}\right).$$

It is easily seen that $K(m)$ is the matrix of Θ with respect to the basis $1, x, \dots, x^{m-1}$. We find that

$$\Theta^2(f(x)) = (-x)^{m-1} f\left(1 - \frac{1}{x}\right)$$

and

$$\Theta^3(f(x)) = (-1)^{m-1} f(x),$$

so $K(m)^3 = (-1)^{m-1} I$, where I is the $m \times m$ identity matrix.

If $m = p^n$, where p is a prime, then we shall show that

$$(-1)^j \binom{m - i - 1}{j} = \binom{i + j - m}{j} \equiv \binom{i + j}{j} \pmod{p} \qquad (*)$$

for $0 \leqslant j < m$. The equality is a simple identity. (Cf. Alan Tucker, *Applied Combinatorics*, Wiley, 1984, p. 210.) To obtain the congruence, observe that for x an indeterminate and k an integer we have the power series identity

$$\begin{aligned}(1 + x)^k &= (1 + x)^{k-m}(1 + x)^{p^n} \\ &\equiv (1 + x)^{k-m}(1 + x^{p^n}) \pmod{p}.\end{aligned}$$

Upon equating coefficients of x^j we obtain

$$\binom{k}{j} \equiv \binom{k - m}{j} \pmod{p},$$

since $j < p^n$. Setting $k = i + j$ gives $(*)$. Thus $K(m) \equiv J(m) \pmod{p}$. Since $(-1)^{m-1} \equiv 1 \pmod{p}$, it follows that $J(m)^3 \equiv I \pmod{p}$.

The solutions given by O. P. Lossers and the proposer involved sums of products of binomial coefficients. Jeffrey Mitchell Cohen showed directly that the matrix $J(m)$ has determinant 1.

The Normalizer Revisited

6528 [1986, 659]. *Proposed by William P. Wardlaw, United States Naval Academy, Annapolis, MD.*

The following is problem 11 on page 44 of C. Chevalley's *Fundamental Concepts of Algebra*, Academic Press, 1956.

Let H be a subgroup of a group G. Show that the elements s of G such that the mapping $t \to sts^{-1}$ of G into itself maps H into itself form a subgroup N of G, of which H is a normal subgroup; show that every subgroup of G containing H and in which H is normal is contained in N.

Provide an example to show that the assertion of the problem is false.

Solution by Aage Bondesen, Søndermarken 25, 3060 *Espergaerde, Denmark.* Let G be the group of permutations of the set $\{1, 2, \dots\}$ of natural numbers, and let H be the subgroup consisting of those elements of G that map every odd number onto

itself. Define

$$N = \{s \in G \mid sHs^{-1} \subset H\}.$$

Then define $s \in G$ by $s(1) = 2$ and

$$s(2m) = 2m + 2, \qquad s(2m+1) = 2m - 1, \quad m = 1, 2, 3, \dots .$$

For every t in H it follows that

$$sts^{-1}(2m-1) = st(2m+1) = s(2m+1) = 2m - 1, \qquad m = 1, 2, 3, \dots .$$

Thus $sts^{-1} \in H$, and, hence, $s \in N$. On the other hand, let $t \in H$ map 2 onto 4 (t could then, for instance, map 4 onto 2, and every other natural number onto itself). Then $s^{-1}ts(1) = s^{-1}t(2) = s^{-1}(4) = 2$, so $s^{-1}ts \notin H$ and $s^{-1} \notin N$. Since $s \in N$ this proves N is not a subgroup. Note also that sHs^{-1} is a proper subgroup of H.

Editorial Comment. A number of solutions used essentially this construction; Bondesen's exposition seemed the most elementary. Many readers found counterexamples in $G = SL(2, R)$, while others used only slightly more esoteric G. Counterexamples can also be constructed using HNN extensions. The original assertion *can* be made correct either (i) by assuming all elements of H are of finite order or (ii) by defining N as the set of all s in G such that

$$sHs^{-1} = H$$

or (iii) by defining N as the set of all s in G such that

$$sHs^{-1} \subseteq H \text{ and } s^{-1}Hs \subseteq H.$$

Also solved by Miroslav D. Ašić, Jeffrey Bergen, Jeffrey Mitchell Cohen, Jesús Ferrer (Spain), Stephen M. Gagola, Jr., Enzo R. Gentile (Argentina), James E. Humphreys, A. A. Jagers (The Netherlands), Charles Lanski, O. P. Lossers (The Netherlands), Pat Morandi, Arthur Rothstein, Eugene Salamin, San Bernardino Problem Solving Group, John Henry Steelman, Douglas B. Tyler, Ken Yanosko, and the proposer.

REVIEWS

EDITED BY JOSEPH KONHAUSER
Department of Mathematics, Macalester College, St. Paul, MN 55105

Episodes in the Mathematics of Medieval Islam. By J. L. Berggren. Springer-Verlag, New York, 1986. xiv + 197 pp., $23.00.

GERALD J. TOOMER
History of Mathematics Department, Brown University, Providence, Rhode Island 02912

Although the author does not say so, it is clear from the book's tone and structure ("exercises" at the end of each chapter) that it is intended to be used primarily as a textbook in undergraduate courses in the history of mathematics. He tells us that "the book reflects my conviction that a proper study of the history of mathematics begins with a study of the texts themselves, and for this reason I have written the chapters not as catalogues of results but rather as expositions of portions of mathematical treatises." So it is pertinent to ask, first, is it useful and practical to present, to an undergraduate audience, the mathematics of earlier epochs in the form of "the texts themselves"; and, secondly, how successful is the present attempt?

As a historian of mathematics, I agree completely with Berggren's emphasis on the importance of reading the original texts. But when one is dealing with mathematical texts from the ancient or medieval periods, the difficulties of presenting them to an undergraduate audience are formidable. Simply translating them is obviously insufficient, since the differences in terminology, notation, and the general approach to mathematical exposition are so great that a "faithful" translation requires explanatory annotations longer than the original to aid the uninitiated, most of whom would quickly be deterred by this approach. Moreover, one has to assume that for the typical undergraduate (whether of a scientific or humanistic orientation), the ancient and medieval periods of human history are a vast series of blank pages occasionally enlivened by entertaining vignettes (such as Archimedes in the bath, to take a "mathematical" example). Yet the cultural background is of real importance in understanding why the practitioners of mathematics were interested in certain topics: the central role of astrology, for instance, is rightly stressed by Berggren in the present book.

Necessarily, then, the writer of such a book has to supply at least a skeletal background of general historical information. Problems of "truth" immediately arise, for almost any general statement about a civilization requires a complex of qualifying statements to make it acceptable to the historian. This problem is intensified for the ancient and medieval periods by the scarcity of the evidence on many subjects, but, as I know by experience, undergraduates in introductory courses are made very uneasy by what they perceive as uncertainty in their textbooks or instructor (while the latter simply regards it as healthy skepticism). So I understand why (although I still find it difficult to approve) Berggren states as bald facts a number of things which are merely disputed hypotheses (to cite just one with which I disagree, that the application of numerical procedures to Greek planetary models is "at least as early as the time of Apollonius of Perga in the third century B.C.")

In the present case, the problems of the proper "historical" approach are further complicated by the difficulties of defining "Islamic" mathematics. One minor dispute is terminological: since "Islam" is primarily a religious term, it seems inappropriate to use it to qualify a science which had very little to do with religion (especially when a number of its practitioners in the period in question were not Muslims). I prefer "Arabic," although that term too requires many qualifications. But, even when we allow "Islamic" to stand as a shorthand word for a cultural complex, we are still faced with the fact that the mathematics (like all the sciences) of that culture are simply a continuation of the Hellenistic Greek tradition. One of the most remarkable features of Islamic civilization was the way in which it took over and continued, in a different language and mostly in a different geographical area, the scientific heritage of antiquity, which was moribund in the contemporary Byzantine Empire, and in so doing breathed new life into it. There are a number of brilliant achievements in Arabic mathematics, but it has to be viewed as the direct continuation of the Greek tradition (and indeed is unintelligible without that background). Thus making "Islamic mathematics" a separate subject of study is artificial. Berggren, a scholar with a notable record of investigation of previously unstudied medieval mathematical texts, is of course well aware of all this, and of necessity allots some space to laying out the ancient Greek background to the topics he treats. But I should have liked to have seen in the book a more forceful presentation of the essential unity of Greek and Arabic mathematics. On the contrary, where possible, he devotes a part of each chapter to "The Islamic Dimension," in which he attempts to show how the topic is related to peculiarly Islamic features of the civilization. These aspects are usually trivial, and in one case that classification is simply inappropriate. In the chapter on geometry, constructions with a "Rusty Compass" are attributed to the "Islamic Dimension." There are indeed some interesting problems in Arabic mathematics involving the use of a compass restricted to a fixed opening, but there is nothing peculiarly Islamic about this, and the basic idea, as usual, is of Greek origin (although, quite by chance, the one known detailed treatment of the subject from the Hellenistic period survives only in the Arabic translation of a Greek work).

How well has the author performed his admittedly difficult task? It is probably impossible to present the necessary historical background in a way which is at the same time accurate, intelligible and useful in the present context, but the effort is a valiant one. My principal objections to the "historical" first chapter are the use of photographs of buildings which illustrate nothing pertinent (apparently on the principle that any illustration is better than none), and exercises involving the *Dictionary of Scientific Biography* (which encourage the beginner's notion that "research" consists of copying information out of reference works). For the rest of the book, the author has the advantage of being able to select "episodes" of intrinsic interest, and for the most part he has selected well. Readers of all mathematical levels will find things to interest and instruct them. Pedagogically, the presentation of the texts is most successful where the author was able to stick most closely to the originals. This is notably so in the chapter on geometry, which is consistently interesting and intelligible (not coincidentally, this is the part of mathematics where methods and notations have changed least in modern times). In the chapter on arithmetic, Berggren has done an excellent job of presenting the numeral systems (both decimal and abjad-sexagesimal) in such a way that the determined student can

actually decipher some of the Arabic texts which are presented in facsimile. Elsewhere, he relies heavily on modern notation, which is probably necessary in the treatment of algebra if the student is not to be completely put off by the Arabic rhetorical exposition. Trigonometry, a usually dismal subject, is enlivened here by the account of al-Bīrūnī's measurement of the earth (where, for once, the anecdotal approach is enlightening). I suspect that the audience to whom this book is primarily addressed will find considerable difficulty in the chapter on spherics, where too much is presented in a short space on a topic which is utterly strange to most undergraduates. The chapter also suffers from containing a "catalogue of results" (which the author had foresworn in his preface). A preferable approach might have been to treat the astrolabe in enough detail that the reader could actually construct one: this leads to numerous illuminating exercises.

The proofreading has been sloppy, and the misprints or slips extend, unfortunately, to numerals which are significant to the course of the exposition. But the instructor may make a virtue of this deficiency by encouraging the students to "reconstruct" the correct proofs or procedures. A particularly good candidate for this "exercise" is the example of division of polynomials on pp. 115–117. It is immediately apparent that there is something wrong, but it will take an astute and determined student to unmask all 13 numerical errors, and in doing so he will certainly gain a thorough understanding of al-Samaw'al's procedure. I have to confess that I did not succeed until I looked at the original Arabic text.

Mathematics and Optimal Form. By Stefan Hildebrandt and Anthony Tromba. Scientific American Books, Inc., New York, 1985. xii + 215 pp., $29.50.

Frank Morgan
Department of Mathematics, Williams College, Williamstown, Massachusetts 01267

Why is a soap bubble round? That is the question I like to ask new acquaintances who wonder about mathematics. After some persistent encouragement, a good number come up with this excellent answer:

> "Since the pressure is the same throughout the bubble, the curvature must be the same everywhere."

Well, that answer deserves commendation. But does it actually follow that a bubble must be a round sphere? Is the round sphere the only closed smooth surface of constant (mean) curvature? In 1958, A.D. Alexandrov [1] proved that if the surface does not cross or intersect itself (as soap films do not), then it must indeed be a round sphere. Good. But now suppose we would allow self-intersections. In that case, under the additional assumption that the surface is a (possibly deformed) sphere, H. Hopf [6] was still able to prove that it actually must be round. Of course, that left an outside possibility of a closed, constant-mean-curvature surface of another type. In 1986, Henry Wente [10] of the University of Toledo actually found one: a constant-mean-curvature, self-intersecting torus (see Figures 1 and 2).

The computer graphics in this review were kindly provided by that famous team at the University of Massachusetts at Amherst: Professors David Hoffman and Joel

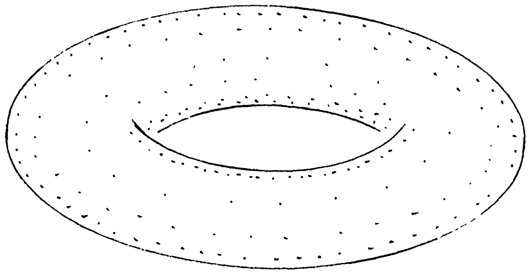


FIG. 1. A torus, but not with constant mean curvature.

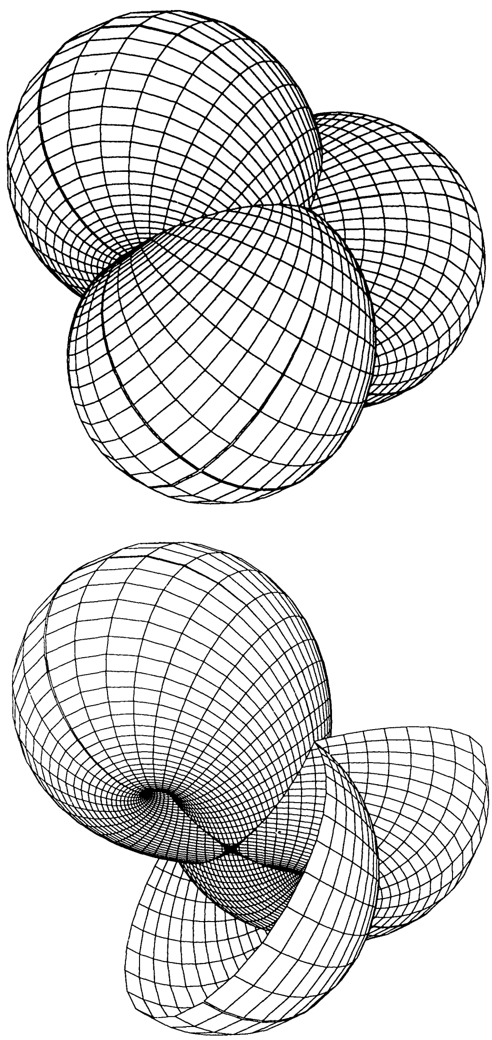


FIG. 2. Wente's twisting constant-mean-curvature torus. Courtesy D. Hoffman, J. Spruck, J. Hoffman, and M. Callahan.

Spruck, graphics programmer Jim Hoffman (no relation), and high school senior Michael Callahan (now at Harvard).

Even more recently, Nicolaos Kapouleas [7], a graduate student at the University of California at San Diego (and now Stanford), produced more closed, constant-mean-curvature surfaces of infinitely many types (of any genus greater than two). They can be approximated by piecing together spheres and the undulating Delaunay surfaces, as suggested by FIGURE 3.

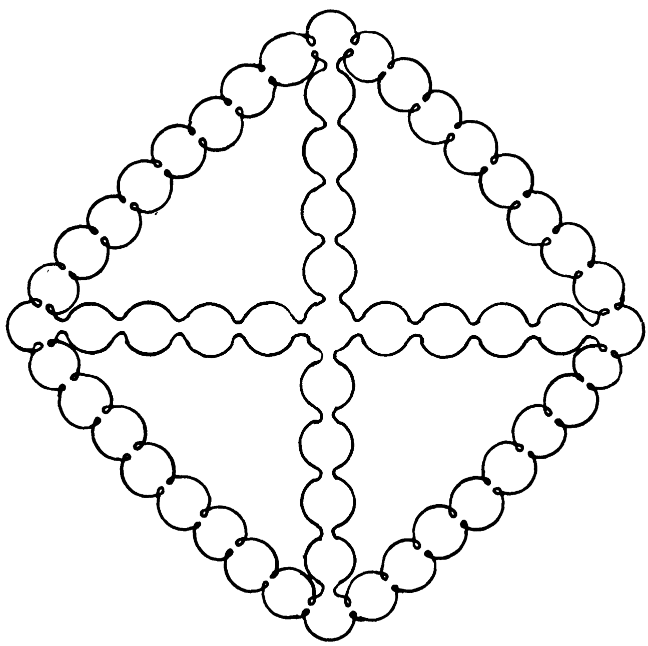


FIG. 3. One of Kapouleas' constant-mean-curvature surfaces (schematic).

Compound bubbles (see FIGURE 4) have more geometry than single ones. In the double bubble the larger bubble, the smaller bubble, and the separating surface all come together along a curve, always at 120° angles. In the triple bubble, four such singular curves meet at a point, always at the "tetrahedral" angle of about 109°. Remarkably, in more complicated froths of thousands of bubbles, no further types of singularities occur. Surfaces meet along curves only three at a time and always at 120° angles. Such curves meet only four at a time and always at about 109° angles. The Belgian physicist, J. A. F. Plateau, observed and recorded these rules over a hundred years ago. But it was not until 1976 that Jean Taylor, now Professor of Mathematics at Rutgers University, succeeded in deducing them from a single principle of area-minimization alone ([9], nicely explained with illustrations in [2]).

Taylor's purely theoretical reasoning stands on its own. One might wonder how she ever came up with the complicated mathematical surfaces she needed. Once when visiting her at home I saw the answer: models of experimental soap films.

These soap bubble singularities are the same singularities which occur in insect wings and radiolarian skeletons.

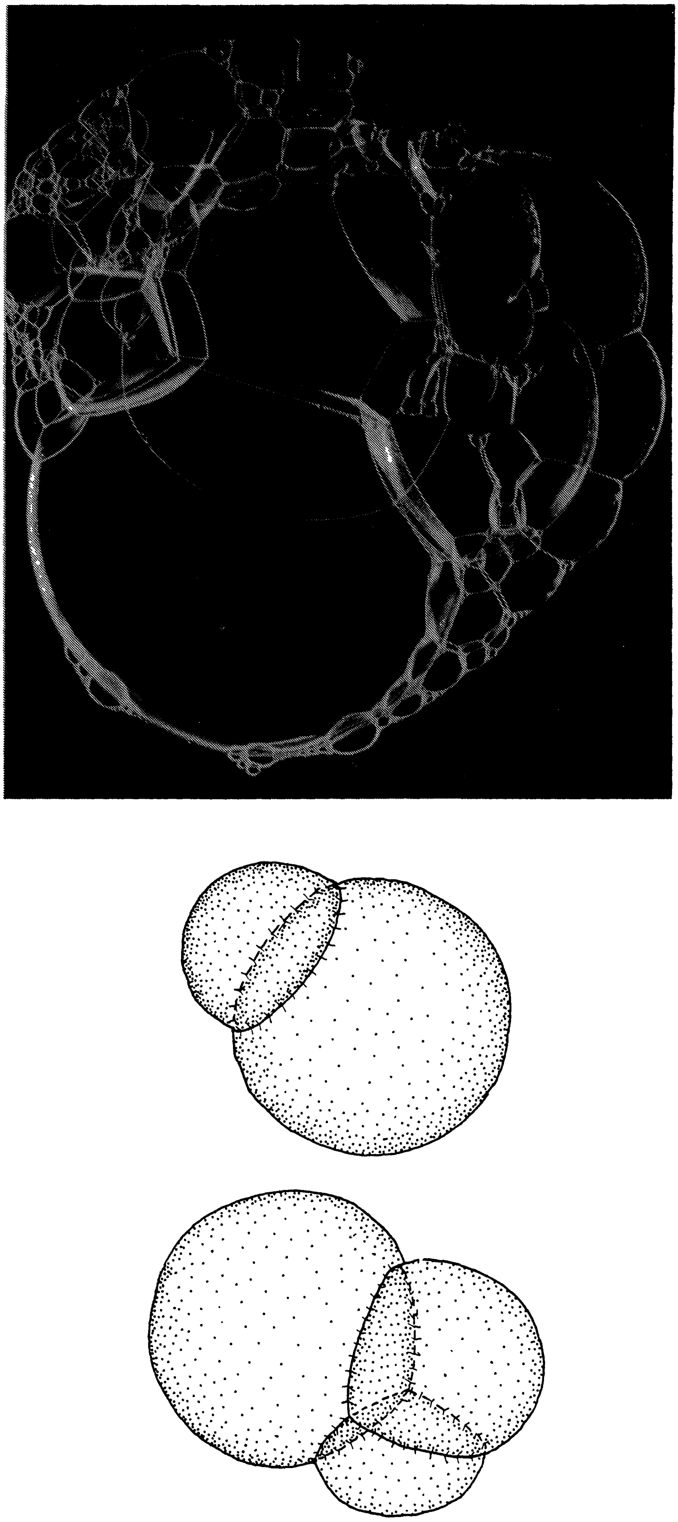


FIG. 4. Compound soap bubbles. Drawings by Jim Bredt. Photograph courtesy of Gordon Gahan/Prism.

In contrast to the unique possible round shape for a single soap bubble, some physical geometric problems have more than one solution. A wire dipped in soap solution may support two different soap films (see FIGURE 5).

FIG. 5. Two soap films with the same boundary. Drawings by Jim Bredt.

It is today an open question whether a smooth boundary can support infinitely many stable soap films. (For an illustrated expository article see [**8**].)

Alternatively consider *complete embedded minimal surfaces*, that is, soap-film-like surfaces that continue on indefinitely, without edge or boundary or self-intersection. Until a few years ago, only two of finite total curvature were known: the plane and the catenoid (see FIGURE 6).

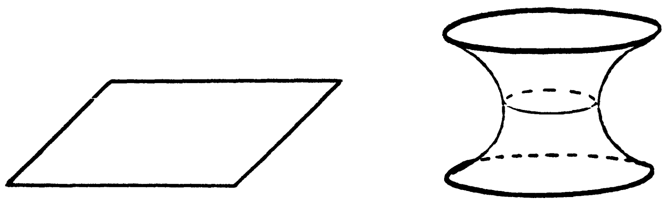


FIG. 6. The plane and the catenoid: the only known complete embedded minimal surfaces of finite total curvature until 1985.

In the past few years, University of Massachusetts Professors David Hoffman and Bill Meeks ([**4**] and [**5**], nicely described and illustrated in [**3**]), have discovered infinitely many others. Their beautiful pictures have adorned the covers of magazines and the walls of offices (see FIGURE 7).

Mathematics and Optimal Form by S. Hildebrandt and A. Tromba provides a magnificent, liberal essay on physical and geometric shapes. Hundreds of beautiful illustrations range from an electron micrograph of a radiolarian skeleton to a rich color reproduction of Chardin's famous painting of the soap-bubble-blower, from brilliant photographs of crystals and liquids in free fall to computer graphics of theoretical soap films. The first simple figure of a Möbius strip is supplemented by a more imaginative sketch by Escher and by an old photograph of A. F. Möbius himself. You may have to read the book to appreciate how the story of the "isoperimetric problem" is illustrated by a medieval map of Cologne and a seventh-century painting of Dido and Aeneas.

A vivid account of Archimedes includes his mathematics and his marvelous contributions to the defense of Syracuse against the Romans. In some interesting

FIG. 7. One of the new complete embedded minimal surfaces discovered by D. Hoffman and W. Meeks. (Courtesy D. Hoffman, J. Spruck, J. Hoffman, and M. Callahan)

selections taken from Plutarch, the Roman commander Marcellus refers to Archimedes as "this geometric Briareus [a hundred-armed giant of Greek legend]."

Gradually the book proceeds to lead the general reader enchantingly from ancient Greece right into modern mathematics. Unfortunately, after giving such a convincing and sympathetic portrayal of Archimedes, it generally omits even the names of contemporary mathematicians. This omission seems especially sad and ironic in this kind of book, so long awaited, which explains the mathematician so well.

From the beginning to end the book provides readable mathematics and applications, colorful history, and even art and poetry. The opening pages give an alluring new translation of Morgenstern's poem, "The Bewitchment," about the magic of blowing bubbles.

Anyone who likes to read and think should love this book. As a mathematician, I want to give it to all my friends.

REFERENCES

1. A. D. Alexandrov, Uniqueness theorems for surfaces in the large, V. Vestnik, Leningrad Univ. 13, No. 19 (1958), 5–8, *Amer. Math. Soc. Trans.*, (2) 21, 412–416.
2. F. J. Almgren and J. E. Taylor, Geometry of soap films, *Scientific American* (July 1976).
3. David Hoffman, The computer-aided discovery of new embedded minimal surfaces, *Math. Intelligencer*, 9 (1987) 8–21.
4. David Hoffman and William Meeks, A complete embedded minimal surface in R^3 with genus one and three ends, *J. Diff. Geom.*, 21 (1985) 109–127.
5. David Hoffman and William Meeks, Complete embedded minimal surfaces of finite total curvature, *Bull. Amer. Math. Soc.*, 12 (1985) 134–136.

6. Heinz Hopf, Differential Geometry in the Large (Seminar Lectures New York Univ. 1946 and Stanford Univ. 1956), Lecture Notes in Math. 1000, Springer-Verlag, 1983.
7. Nicolaos Kapouleas, Compact constant mean curvature surfaces in Euclidean three space, preprint. cf. *Bull. Amer. Math. Soc.*, 17 (1987) 318–20.
8. Frank Morgan, Soap films and problems without unique solutions, *Amer. Scientist* (May, 1986) 232–236.
9. J. E. Taylor, The structure of singularities in soap-bubble-like and soap-film-like minimal surfaces, *Annals of Math.*, 103 (1976) 489–539.
10. Henry C. Wente, Counterexample to a conjecture of H. Hopf, *Pacific J. Math.*, 121 (1986) 193–243.

TELEGRAPHIC REVIEWS

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T: Textbook	P: Professional Reading	1–4: Semesters
C: Computer Software	L: Undergraduate Library	**: Special Emphasis
S: Supplementary Reading	13: Grade Level	??: Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

General, S*(18), P, L*.** *Proceedings of the International Congress of Mathematicians, August 3-11, 1986.* Ed: Andrew M. Gleason. AMS, 1987, $195 set [ISBN: 0-8218-0110-4]. *Volume 1,* ci + 872 pp; *Volume 2,* iii + 837 pp. Well-written expository articles by prominent researchers on central questions in mathematics. The two volumes contain reports on the work of the Fields medalists and Nevanlinna prize winner, fifteen plenary addresses, and over one hundred invited addresses grouped by field, as well as a list of short communications. Focusing on the idea rather than the technical, the articles successfully give the mathematician an appreciation of topics outside his or her own specialty. The proceedings also include coverage of the opening and closing ceremonies, and several lists of committees, donors, and participants. GG

Elementary, T*((13: 1). *Mathematics for Liberal Arts: A Problem Solving Approach.* Rick Billstein, Johnny W. Lott. Benjamin/Cummings, 1986, xvii + 646 pp, $29.95. [ISBN: 0-8053-0863-6] A book covering basic algebra, but introducing it in such a way as to emphasize the structure behind it. After covering the basics, it uses them in probability, statistics, and linear programming. MZ

Education, P*, L.** *Science Achievement in Seventeen Countries: A Preliminary Report.* IEA. Pergamon Pr, 1988, xi + 125 pp, (P). [ISBN: 0-08-036563-9] First report of an international assessment of science education for students of ages 10, 14, and 18. U.S. results in science are comparable to those in mathematics (see *The Underachieving Curriculum,* TR, March 1987): very low. U.S. performance declined as children got older, from middle for 10-year-olds to very low (last in biology) for students who are still studying science as high school seniors. One sub-study reports that the bottom 25% of U.S. 14-year-olds perform only at chance level. LAS

Education, P, L. *Perspectives on Research on Effective Mathematics Teaching, Volume 1.* Ed: Douglas A. Grouws, Thomas J. Cooney, Douglas Jones. NCTM, 1988, viii + 261 pp, $15 (P). [ISBN: 0-87353-254-6] One of a series of four reports from NCTM-sponsored conferences intended to establish research agendas in areas of mathematics education where conceptual consensus is feasible. Twelve survey papers outline issues in teaching amenable to research, including higher-order thinking, cross-cultural studies, content determination, knowledge construction, and professionalism of teachers. Proceedings of other conferences—on problem solving, on learning algebra, and on number concepts—will be published subsequently, as will an overview volume. LAS

History, L*, P. *I Have A Photographic Memory.* Paul R. Halmos. AMS, 1987, vii + 326 pp, $55. [ISBN: 0-8218-0115-5] Six-hundred snapshots of mathematicians culled from Halmos's collection of 6000, captioned with chatty, informative commentary on mathematics, meetings, and careers. Arranged more or less chronologically from the early 50's to the mid 80's. Index of names. A unique, visual portrait of the mathematical community. LAS

Logic, T*(15-17: 1, 2), L*. *Computability and Logic.* Daniel E. Cohen. Math. & Its Applic. Halsted Pr, 1987, 243 pp, $64.95. [ISBN: 0-470-20847-3] Textbook for introductory course in logic or computability. First part covers elementary recursion theory. Includes chapter on Hilbert's tenth problem. Second part develops propositional and predicate calculus through completeness theorems using natural deduction. Final two chapters cover incompleteness and undecidability theorems and decidability of natural numbers under addition. KS

Linear Algebra, T*(13-14: 1, 2), S, L. *A Unified Introduction to Linear Algebra: Models, Methods, and Theory.* Alan Tucker. Macmillan, 1988, xiii + 540 pp. [ISBN: 0-02-421580-5] An introduction to matrix algebra via extensive models that illus-

trate important issues in applied linear algebra: calculation, sensitivity, appropriateness of linearity, estimation, errors, etc. Applications include digital filtering, Markov processes, linear programming, adjacency matrices, condition numbers, and all canonical examples. The text consists predominantly of page-long examples of models, loosely glued together with definitions, occasional theorems, and boxed rules of calculation. Good exercises, both routine and innovative. An imaginative, stimulating text for "the most useful college mathematics course most students ever take." LAS

Calculus, T(13: 2-3). *Calculus.* William E. Boyce, Richard C. DiPrima. Wiley, 1988, xvii + 1124 pp, $54.82. [ISBN: 0-471-09333-5] Fairly standard set of topics. Good applications, mostly from mechanics and physics. Numerical methods are reasonably treated. Odd numbered answers included. Solution manual available. MR

Calculus, T(13: 2). *Calculus and Analytic Geometry, Third Edition.* Philip Gillett. DC Heath, xix + 1034 pp, $38. [ISBN: 0-669-13100-8] Major changes include addition of new exercises, moving the trigonometry review to Chapter 1, having limits precede derivatives, adding a new section on differentiation of sines and cosines, and moving limits involving infinity to Chapter 5. (*First Edition,* TR, August-September 1981.) LC

Calculus, T(17: 2), S. *An Introduction to the Calculus of Variations.* Charles Fox. Dover, 1987, viii + 271 pp, $6.95 (P). [ISBN: 0-486-65499-0] Covers all the important techniques including first and second variation, principle of least action, and Weierstrassian E-function. Many worked examples and references. Some exercises with no solutions included. Would serve as a good reference but not as a text for an undergraduate course. MR

Calculus, T(13: 1). *Calculus, Fourth Edition.* Marvin L. Bittinger. Addison-Wesley, 1988, xviii + 570 pp, $39.75. [ISBN: 0-201-12216-2] For a one-semester intuitive calculus course for non-mathematics majors. "Applications" in a wide variety of disciplines. Loaded with pedagogical features. Supplementary materials available include an instructor's manual, a test bank, videotapes, computerized testing, computer software, and a student's solutions manual. What is left for the teacher? (*First Edition,* TR, June-July 1976; *Second Edition,* TR, January 1981; *Third Edition,* TR, August-September 1985.) JK

Calculus, T(13-14: 3). *Calculus and Analytic Geometry, Seventh Edition.* George B. Thomas, Jr., Ross L. Finney. Addison-Wesley, 1987, xv + 1253 pp, $48.95. [ISBN: 0-201-16320-9] Now contains chapter reviews, historical notes, lessons in three-dimensional drawing, enhanced artwork, and quick-reference charts that group formulas, emphasize procedures, and describe problem-solving strategies. Among many revisions is a considerably changed presentation of multivariable calculus. Many new applications. Same level of rigor. (*Fourth Edition,* TR, December 1968; Extended Review, June-July 1970; *Alternate Edition,* TR, June-July 1973; *Fifth Edition,* TR, June-July 1979.) DFA

Calculus, T(13: 2). *Calculus with Analytic Geometry, Second Alternate Edition.* Earl W. Swokowski. Prindle, Weber & Schmidt, 1988, xiii + 914 pp. [ISBN: 0-87150-008-6] Contains essentially the same material as *Calculus with Analytic Geometry, Fourth Edition,* but this *Alternate Edition* introduces trigonometric functions earlier. (*First Edition,* TR, December 1975; *Second Edition,* TR, June-July 1979; *Third Edition,* TR, November 1984.) JNC

Calculus, T(13: 2). *Calculus with Analytic Geometry, Alternate Edition.* Robert Ellis, Denny Gulick. Harcourt Brace Jovanovich, 1988, xi + 1069 pp, $39. [ISBN: 0-15-505700-6] The major difference between this and the *Third Edition* is that here the chapter on applications precedes rather than follows techniques of integration. Other changes include introducing limits via tangent lines and velocity, and a new section on integration using tables. (*First Edition,* TR, August-September 1978; *Third Edition,* TR, October 1986.) JNC

Calculus, T(13: 2). *Calculus and Analytic Geometry, Alternate Second Editon.* C.H. Edwards, Jr., David E. Penney. Prentice-Hall, 1988, xiii + 1089 pp. [ISBN: 0-13-111469-7] This *Alternate Edition* has an earlier introduction of the Chain Rule, later coverage of trigonometric functions, and a more intuitive introduction to exponential and logarithmic functions. (*First Edition,* TR, October 1983; *Second Edition,* TR, May 1986.) JNC

Calculus, T(13-14: 3). *Calculus and Analytic Geometry, Fourth Edition.* Al Shenk. Scott Foresman, 1988, 1135 pp, $35.96. [ISBN: 0-673-16721-6] Extensively rewritten edition, now at a higher level; primarily for mathematics, science, and engineering students. Some notation and terminology have now been standardized. Rearranged presentation of many topics, following reader suggestions. New section on Jacobians and changing variables in multiple integrals. Manuals, computer supplements, transparencies, test bank available. DFA

Calculus, T(13: 1, 2). *Mathematical Analysis for Business and Economics, Second Edition.* Charles W. Schelin, David W. Bange. Ser. in Math. Prindle, Weber & Schmidt, 1988, xiii + 601 pp. [ISBN: 0-534-91493-4] Covers mathematical topics of use to students in business and economics, including calculus, multivariate models, sequences and series, probability, matrix algebra, linear programming, and curve fitting. The topics are introduced through examples. The mathematical concepts are drawn from the examples and then developed and applied to more general situations. Applications to business and economics are strongly emphasized in the examples and exercises. A good background in algebra is assumed, but a review is provided in an appendix. RH

Calculus, T(13: 1, 2). *Applied Calculus, Second Edition.* Marvin L. Bittinger, Bernard B. Morrel. Addison-Wesley, 1988, xix + 779 pp, $41.95. [ISBN:

0-201-12211-1] Many varied examples and problems for students in business and the social sciences. Standard chapters on differentiation; delays treatment of trigonometry. Integration techniques are substitution, parts, and tables. Nice chapters on differential equations and numerical methods. Several variables chapter includes least squares and Lagrange multipliers, but applies only local methods to global extremum problems! Error estimates for Taylor series are not discussed! Features include exercises in the margins, answers to odd problems, and formulas on the book's covers. A test generator, graphing software, and student solution manual are available. GG

Calculus, T(13: 2). *Calculus: Theory and Applications in Technology and the Physical and Life Sciences.* R.M. Johnson. Math. & Its Applic. Halsted Pr, 1987, 333 pp, $39.95 (P). [ISBN: 0-470-20898-8] An expensive, compact paperback whose orientation reflects its origin in a British College of Technology. Covers the usual calculus topics but minimizes the motivational discussion and number of examples. JNC

Complex Analysis, T(18: 1), P. *Applications of Harmonic Measure.* John B. Garnett. Lect. Notes in Math. Sci., V. 8. Wiley, 1987, vii + 69 pp, $22.95 (P). [ISBN: 0-471-62772-0] After a concise but lucid introduction to harmonic measure, the author presents several applications to such topics as level curves, interpolating sequences, and periodic spectra of Hill's equation. No exercises, good bibliography. TAV

Differential Equations, T(15-17: 1), S, P, L. *Mathematics for Dynamic Modeling.* Edward Beltrami. Academic Pr, 1987, xvi + 277 pp, $27.50. [ISBN: 0-12-085555-0] A text for an applications-oriented second course in differential equations, as taught by non-pure mathematics departments. "A number of central results are motivated and explained by a combination of a little rigor and a lot of intuition." Equilibrium and stability in differential equations modeling; orbit behavior of nonlinear models. Dozens of exercises. RB

Partial Differential Equations, T(17: 1).** *First-Order Partial Differential Equations, Volume I: Theory and Application of Single Equations.* Hyun Ku Rhee, Rutherford Aris, Neal R. Amundson. Intern. Ser. in Physical & Chemical Eng. Sci. Prentice-Hall, 1986, xiii + 543 pp, $45.95. [ISBN: 0-13-319195-8] A very nice book on partial differential equations intended to be used by first-year engineering students, so it has many applications of partial differential equations. As well as applications, it contains the theory of how to arrive at the equations and the different techniques needed to solve them. MZ

Partial Differential Equations, P. *Nonlinear Elliptic and Parabolic Equations of the Second Order.* N.V. Krylov. Math. & Its Applic. Kluwer Academic, 1987, xiii + 462 pp, $99. [ISBN: 90-277-2289-7] The author constructs a solvability theory "covering the Bellman equation, the Monge-Ampère equation, and the quasilinear equations under natural conditions on their coefficients," based on differential equations theory without the usual recourse to probabilistic or geometric methods. RB

Partial Differential Equations, P. *Recent Topics in Nonlinear PDE II.* Ed: Kyûya Masuda, Masayasu Mimura. Math. Stud., V. 128. Elsevier Science, 1985, vii + 228 pp, $70 (P). [ISBN: 0-444-87938-2] A collection of ten papers based on lectures given at a meeting on nonlinear partial differential equations given at Tohoku University, February 27-29, 1984. Topics presented range over various areas of mathematical physics. AM

Numerical Analysis, P. *Difference Schemes: An Introduction to the Underlying Theory.* S.K. Godunov, V.S. Ryabenkii. Transl: E.M. Gelbard. Stud. in Math. & Its Applic., V. 19. Elsevier Science, 1987, xvii + 489 pp, $77.75. [ISBN: 0-444-70233-4] Translation of the 1977 Russian edition. Convergence and stability theory of various difference schemes that arise in the numerical solution of the ordinary differential equations and partial differential equations of boundary value problems. RWN

Analysis, T(18: 2), P. *Ergodic Theory and Differentiable Dynamics.* Ricardo Mañé. Transl: Silvio Levy. Ergebnisse der Math. und ihrer Grenzgebiete 3 Folge, Band 8. Springer-Verlag, 1987, xii + 317 pp, $82. [ISBN: 0-387-15278-4] A translation from the Portuguese edition. Following a quick review of measure theory, the author presents an overview of the nature of dynamical systems that generate fundamental questions of ergodicity. The following two sections on ergodicity and entropy are modern in approach and quite deep. Many useful exercises and an excellent bibliography. TAV

Algebraic Geometry, P. *Lecture Notes in Mathematics-1273: Singularities, Representation of Algebras, and Vector Bundles.* Ed: G.-M. Greuel, G. Trautmann. Springer-Verlag, 1987, xiv + 383 pp, $32.90 (P). [ISBN: 0-387-18263-2] Proceedings of a symposium held in Lambrecht, West Germany, December 1985, which drew together researchers in diverse topics. Three survey talks and eighteen research articles (half on singularities) by authors from Germany, Norway, India, Holland, United States, and France. RB

Algebraic Geometry, T*(18: 3), P*. *Relative Invariants of Sheaves.* A. Verschoren. Pure & Appl. Math., V. 104. Dekker, 1987, xi + 249 pp. [ISBN: 0-8247-7734-4] Rich in examples and explanation, this introduction to sheaf theory combines methods from relative invariants of rings with classical techniques from sheaf theory. The author derives extension properties of coherent and quasicoherent sheaves, revealing some connections to local cohomology. Includes two treatments of global invariants together with torsion couples and morphisms between them; Krull schemes; Hecke actions; contravariant maps between relative invariants. Beautifully written and filled with insights. There are, however, too few exercises, which may make it difficult to use this book as a text. LW

Differential Geometry, P. *Lecture Notes in Mathematics-1251: Differential Geometric Methods in Mathematical Physics.* Ed: P.L. García, A. Pérez-Rendón. Springer-Verlag, 1987, vii + 300 pp, $31.50 (P). [ISBN: 0-387-17816-3] Proceedings of the fourteenth international conference held June 24-29, 1985 at the University of Salamanca. A collection of fourteen papers covering superalgebras and supermanifolds, superfield theory, classical and quantized field theory, and differential geometric techniques in general. AM

Differential Geometry, P. *Differential Geometry and Its Applications.* Ed: D. Krupka, A. Švec. Math. & Its Applic. D Reidel (US Distr: Kluwer Academic), 1987, xiii + 381 pp, $98. [ISBN: 90-277-2487-3] Proceedings of a conference on the topic at Brno, Czechoslovakia, August 1986. Thirty-four papers in differential geometry and geometric methods in physics, predominantly by eastern European authors. MU

Differential Geometry, T*(17-18), S. *Riemannian Geometry.* S. Gallot, D. Hulin, J. Lafontaine. Universitext. Springer-Verlag, 1987, xi + 248 pp, $29 (P). [ISBN: 0-387-17923-2] A graduate text which starts with abstract manifolds rather than the local theory of embedded surfaces, and which illustrates new concepts and results with a repertoire of recurrent examples (including spheres, projective spaces, compact Lie groups). Presumes some exposure to elementary notions of differential manifolds, representation theory, homotopy. Hundreds of *solved* exercises! Intriguing approach. RB

Differential Geometry, P. *Seminar on New Results in Nonlinear Partial Differential Equations.* Anthony J. Tromba. Aspects of Math., V. E10. Friedr Vieweg & Sohn, 1987, vi + 198 pp, (P). [ISBN: 3-528-08975-X] A collection of eight papers describing results reported during the "year" in partial differential equations, January-September 1984 at the Max-Planck-Institut für Mathematik. Participants included Jerry Kazdan, Tudor Ratiu, Jerry Marsden, Henry Wente, Anthony Tromba, Wei-Ming Ni, Vincent Moncrief, Uwe Abresch, and Sergiu Klainerman. AM

Algebraic Topology, P. *Algebraic Topology and Algebraic K-Theory.* Ed: William Browder. Annals of Math. Stud., No. 113. Princeton U Pr, 1987, ix + 563 pp, $65; $22.50 (P). [ISBN: 0-691-08415-7; 0-691-08426-2] Proceedings of a Princeton conference, October 1983, in honor of the sixtieth birthday of John C. Moore, the "doctor-father" of numerous prominent topologists, including Swan, Browder, Stasheff, and May. Twenty papers on homotopy theory, algebraic topology more generally, and aspects of K-theory related to topology. RB

Operations Research, T(16-18: 1, 2), P. *Mathematical Programming: Theory and Algorithms.* M. Minoux. Transl: Steven Vajda. Wiley, 1986, xxviii + 489 pp, $66.95. [ISBN: 0-471-90170-9] A survey of mathematical programming algorithms. Presents a unified introduction to linear programming, constrained and unconstrained nonlinear programming, integer programming, and dynamic programming. Each chapter includes an extensive bibliography. AO

Probability, P. *Lecture Notes in Mathematics-1193: Geometrical and Statistical Aspects of Probability in Banach Spaces.* Ed: X. Fernique, *et al.* Springer-Verlag, 1986, 128 pp, $11.60 (P). [ISBN: 0-387-16487-1] Proceedings of a conference held in Strasbourg, 1985. Starts with a brief survey of work of Antoine Ehrhard. Covers invariance principles for the empirical measure of a mixing sequence, and for the local time of Markov processes; almost exchangeable sequences in $L^q, 1 \leq q < 2$; uniform convergence of Gaussian and Rademacher Fourier quadratic forms; metric entropy and the central limit theorem in Banach spaces. MZ

Probability, P. *Lecture Notes in Mathematics-1206: Probability and Analysis.* Ed: G. Letta, M. Pratelli. Springer-Verlag, 1986, viii + 283 pp, $23.40 (P). [ISBN: 0-387-16787-0] Lectures given at the First 1985 Session of the Centro Internazionale Matematico Estivo held at Varenna, Italy. Lectures cover such topics as probability and geometry, Martingales and Fourier analysis, cylinder measures, and local bases and nuclearity. MZ

Stochastic Processes, T(18: 1), P. *Lectures on Stochastic Analysis: Diffusion Theory.* Daniel W. Stroock. Math. Soc. Stud. Texts, V. 6. Cambridge U Pr, 1987, ix + 128 pp, $11.95 (P); $34.50. [ISBN: 0-521-33645-7; 0-521-33366-0] After an introductory section on Wiener measure and its setting, the book contains an introduction to diffusion theory, Martingale theory, and the Ito calculus. The tract concludes with the formulation of the Martingale problem of diffusion. Quite technical. Surprisingly, no bibliography. TAV

Stochastic Processes, P. *Ergodic Theory of Random Transformations.* Yuri Kifer. Progress in Prob. & Stat., V. 10. Birkhauser Boston, 1986, 210 pp, $34. [ISBN: 0-8176-3319-7] A study of ergodic theoretical properties of evaluation processes generated by independent application of transformation chosen according to some probability distribution. Applications to problems of dynamical systems and stochastic flows. TAV

Statistics, T(15-18: 1, 2), S, L. *Statistical Methods for the Analysis of Biomedical Data.* Robert F. Woolson. Prob. & Stat. Wiley, 1987, xx + 513 pp, $49.95. [ISBN: 0-471-80615-3] Reference guide or text for biostatistics. Presupposes only algebra. Describes procedures by presenting worked examples. Chapters on epidemiological and clinical data and on survival curves. FLW

Statistics, T(17: 1), S. *Multiple Comparison Procedures.* Yosef Hochberg, Ajit C. Tamhane. Prob. & Math. Stat. Wiley, 1987, xxii + 450 pp, $44.95. [ISBN: 0-471-82222-1] This textbook addresses many controversial subjects connected with multiple comparison techniques including their misuse in practice. It includes an excellent bibliography of relevant literature. Includes detailed derivations and operat-

ing characteristics for procedures, as well as examples for implementation. The first part deals with fixed effects models, while the second part deals with other models and alternative approaches. Good for researchers and practitioners. LB-E

Programming, T(14-18), S, P, L. *Introduction to Common Lisp.* Taiichi Yuasa, Masami Hagiya. Transl: Richard Weyhrauch, Yasuko Kitajima. Academic Pr, 1987, ix + 293 pp, $29.95. [ISBN: 0-12-774860-1] Common Lisp is a completely specified, hardware and operating system-independent dialect of the programming language Lisp; the authors produced one of the first implementations (Kyoto Common Lisp). Here they present Common Lisp for Lisp beginners (with computer science experience) and a Common Lisp reference for Lisp programmers. Examples throughout, including complete programs. RB

Programming, T*(13), S, P*, L*.** *The Little LISPer, Trade Edition.* Daniel P. Friedman, Matthias Felleisen. MIT Pr, 1987, xiv + 186 pp, $12.95 (P). [ISBN: 0-262-56038-0] An extension of a noted text for novices which quickly introduces recursive thinking via Lisp to novices. Asserting that Lisp programming is essentially pattern recognition, these Indiana University authors instruct by means of an inviting, informal question-answer format, allowing combined advantages of abundant answered exercises and conversational tone. Aptly suited to non-programmers without quantitative backgrounds. (*First Edition,* TR, January 1975.) RB

Computer Science, P. *Proceedings of the 1987 International Conference on Parallel Processsing.* Ed: Sartaj K. Sahni. Penn St U Pr (215 Wagner Building, University Park, PA 16802), 1987, xx + 923 pp, $85 (P). [ISBN: 0-271-00608-0] Over 170 papers (acceptances were only 36% of submission): architecture, algorithms, compiler techniques, interconnection networks, hypercube computing, software support, image processing, logic programming and systems, systolic computing, numerical algorithms, multiprocessor concurrency, data structures, graph algorithms, programming languages, fault tolerance and testing, programming systems, embeddings. RB

Applications (Biological Science), S(18), P, L. *Lecture Notes in Biomathematics-72: Modeling and Management of Resources under Uncertainty.* Ed: T.L. Vincent, *et al.* Springer-Verlag, 1987, viii + 318 pp, $32.60 (P). [ISBN: 0-387-17999-2] Proceedings of the Second U.S.-Australia Workshop on Renewable Resource Management held in Honolulu in December 1988. Divided into three parts: modeling/biology, control/techniques, management/real problems. Twenty-three papers plus referees' comments. SM

Applications (Fluid Mechanics), T(18). *Topics in Geophysical Fluid Dynamics: Atmospheric Dynamics, Dynamo Theory, and Climate Dynamics.* M. Ghil, S. Childress. Appl. Math. Sci., V. 60. Springer-Verlag, 1987, xv + 485 pp, $39 (P). [ISBN: 0-387-96475-4] The book is divided into four parts, starting with fundamentals like the effects of rotation, shallowness, and the quasi-geostrophic approximation. The rest covers large-scale atmospheric dynamics, the dynamo theory, and theoretical climate dynamics. MZ

Applications (Physics), P. *General Relativity and Gravitation.* Ed: M.A.H. MacCallum. Cambridge U Pr, 1987, xvii + 407 pp, $49.50. [ISBN: 0-521-33296-6] Proceedings of the eleventh international conference held in Stockholm, July 1986. Thirteen plenary lectures surveyed recent advances: black holes; supergravity, Kaluza-Klein, superstrings; quantum gravity; twistors; numerical relativity, gravitational radiation, cosmology; inflation in the universe. Reports by chairmen on symposia in which over 500 papers and posters were presented. MU

Applications (Physics), P. *Introduction to Superanalysis.* Felix Alexandrovich Berezin. Ed: A.A. Kirillov. Transl: J. Niederle, R. Kotecký. Math. Physics & Appl. Math., V. 9. D Reidel (US Distr: Kluwer Academic), 1987, xii + 424 pp, $109. [ISBN: 90-277-1668-4] The first three chapters, Grassman Algebra, Superanalysis, and Linear Algebra in Z-Graded Spaces, were virtually completed before Berezin's untimely death. The last two chapters, Supermanifolds in General, and Lie Superalgebras, were compiled largely from his preprints and notes. MU

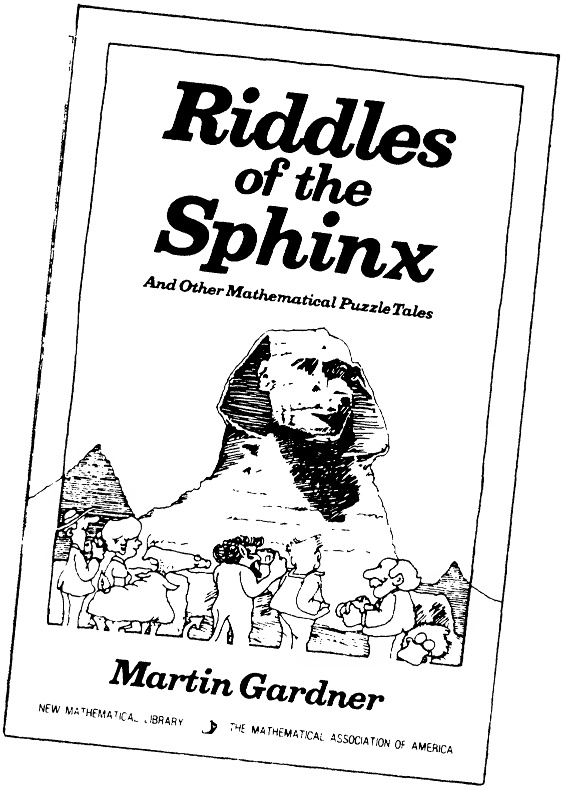
Applications (Physics), T(18: 1), S, P. *The Interacting Boson Model.* F. Iachello, A. Arima. Mono. on Math. Physics. Cambridge U Pr, 1987, x + 250 pp, $59.50. [ISBN: 0-521-30282-X] Using the theory of group transformations, this book describes the mathematical framework on which the interacting boson model of the nucleus is based and discusses applications to the calculation of properties of nuclei. MU

Reviewers

RJA: Richard J. Allen, St. Olaf; DFA: David F. Appleyard, Carleton; RB: Richard Brown, Carleton; LB-E: Lisa Brown-Ewing, Macalester; JNC: Judith N. Cederberg, St. Olaf; LC: Laura Chihara, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; RSF: Robert S. Fisk, St. Olaf; SG: Steven Galovich, Carleton; GG: George Gilbert, St. Olaf; BH: Bruce Hanson, St. Olaf; RH: Rhonda Hatcher, St. Olaf; RK: Roger Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; AM: Alan Magnuson, St. Olaf; SM: Steve McKelvey, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; MR: Matthew Richey, St. Olaf; AWR: A. Wayne Roberts, Macalester; GMS: G. Michael Schneider, Macalester; PS: Paul Schoening, St. Olaf; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MS: Myriam Steinback, Macalester; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; LW: Lynne Walling, St. Olaf; FLW: Frank L. Wolf, Carleton; MZ: Michael Zielinski, St. Olaf.

Riddles of the Sphinx

and other mathematical puzzle tales.



by Martin Gardner
Volume 32 in the New Mathematical Library
184 pp., Paper. ISBN-0-88285-632-8
List: $14.50 MAA Member: $12.50

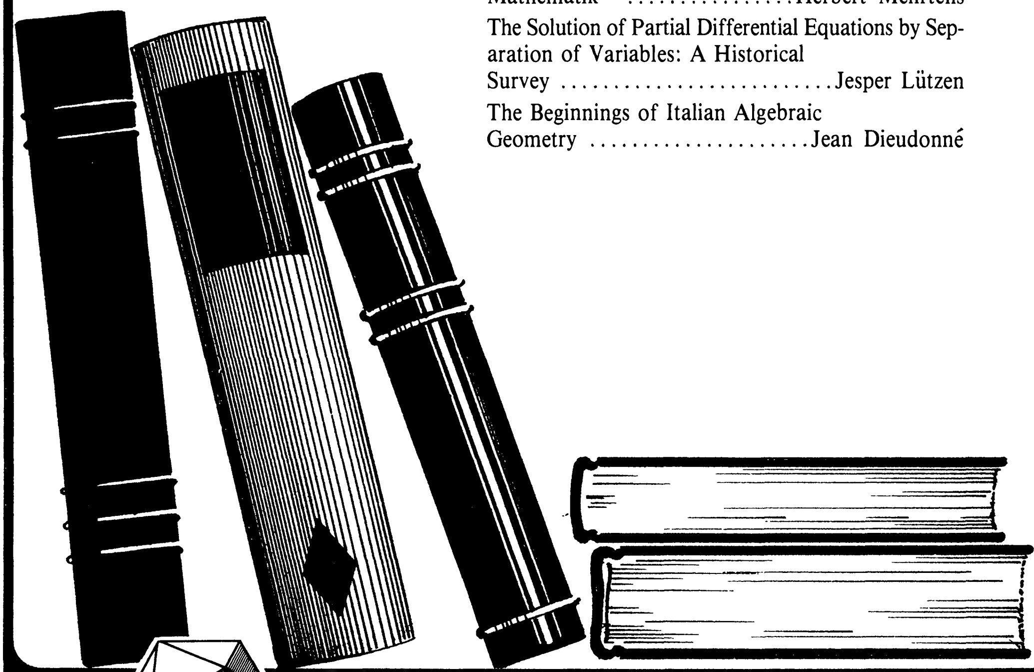
Martin Gardner has charmed readers for over fifty years with his delightful books and articles on science. He is best known for the popular column, *"Mathematical Games,"* which appeared in *Scientific American* for twenty-five years. Generations of scientists and mathematicians have been inspired by his writing and the MAA is proud to include his name in its list of authors.

This book was drawn from Gardner's column in *Isaac Asimov's Science Fiction Magazine.* The riddles presented here incorporate the responses of his initial readers, along with additions suggested by the editors of the New Mathematical Library. Each chapter (riddle) poses a problem answered in the First Answers section. The solution in turn raises another problem that is solved in the Second Answers section. This may suggest a third question and in several instances there is a fourth. Gardner draws us from questions to answers always presenting us with new riddles—some as yet unanswered. There are 125 different pieces altogether.

Solving these riddles is not simply a matter of logic and calculation, although these play a role. Luck and inspiration are factors as well, so beginners and experts alike may profitably exercise their wits on Gardner's problems, whose subjects range from geometry to word play to questions relating to physics and geology.

We guarantee that you will solve some of the riddles, be stumped by others, and be amused by almost all of the stories and settings that Gardner has devised to raise these questions.

Order from:
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036



THE AMERICAN
MATHEMATICAL MONTHLY

Volume 95, Number 7 August–September 1988

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Dept. of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Shorter articles about the teaching of mathematics should be sent to STAN WAGON and JOAN HUTCHINSON, Department of Mathematics, Smith College, Northampton, MA 01063. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to ANITA E. SOLOW, Dept. of Mathematics, Grinnell College, Grinnell, IA 50112; in geometry or topology to DENNIS DETURCK, Dept. of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to DAVID J. HALLENBECK, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Proposed problems (three copies) and solutions (two copies), both elementary and advanced to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, University of Calgary, Alberta, Canada T2N 1N4.

Reprint Permission: A. B. WILLCOX, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Back Issues: P. and H. BLISS Co., Middletown, CT 06457.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $26 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $29. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $90 per year.

Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Assocation of America, 1529 Eighteenth Street, N.W., Washington, DC 20077-9564.

PRINTED IN THE UNITED STATES OF AMERICA

Gauss, Landen, Ramanujan, the Arithmetic-Geometric Mean, Ellipses, π, and the *Ladies Diary*

GERT ALMKVIST, *University of Lund*

BRUCE BERNDT*, *University of Illinois*

GERT ALMKVIST received his Ph.D. at the University of California in 1966 and has been at Lund since 1967. His main interests are algebraic K-theory, invariant theory, and elliptic functions.

BRUCE BERNDT received his A.B. degree from Albion College, Albion, Michigan in 1961 and his Ph.D. from the University of Wisconsin, Madison, in 1966. Since 1977, he has devoted all of his research efforts to proving the hitherto unproven results in Ramanujan's notebooks. His book, *Ramanujan's Notebooks*, Part I (Springer-Verlag, 1985), is the first of either three or four volumes to be published on this project.

Virtue and sense, with female-softness join'd
(All that subdues and captivates mankind!)
In Britain's matchless fair resplendent shine;
They rule Love's empire by a right divine:
Justly their charms the astonished World admires,
Whom Royal Charlotte's bright example fires.

1. Introduction. The arithmetic-geometric mean was first discovered by Lagrange and rediscovered by Gauss a few years later while he was a teenager. However, Gauss's major contributions, including an elegant integral representation, were made about 7–9 years later. The first purpose of this article is, then, to explain the arithmetic-geometric mean and to describe some of its major properties, many of which are due to Gauss.

*Research partially supported by the Vaughn Foundation.

Because of its rapid convergence, the arithmetic-geometric mean has been significantly employed in the past decade in fast machine computation. A second purpose of this article is thus to delineate its role in the computation of π. We emphasize that the arithmetic-geometric mean has much broader applications, e.g., to the calculation of elementary functions such as $\log x$, e^x, $\sin x$, and $\cos x$. The interested reader should further consult the several references cited here, especially Brent's paper **[14]** and the Borweins' book **[13]**.

The determination of the arithmetic-geometric mean is intimately related to the calculation of the perimeter of an ellipse. Since the days of Kepler and Euler, several approximate formulas have been devised to calculate the perimeter. The primary motivation in deriving such approximations was evidently the desire to accurately calculate the elliptical orbits of planets. A third purpose of this article is thus to describe the connections between the arithmetic-geometric mean and the perimeter of an ellipse, and to survey many of the approximate formulas that have been given in the literature. The most accurate of these is due to Ramanujan, who also found some extraordinarily unusual and exotic approximations to elliptical perimeters. The latter results are found in his notebooks and have never been published, and so we shall pay particular attention to these approximations.

Also contributing to this circle of ideas is the English mathematician John Landen. In the study of both the arithmetic-geometric mean and the determination of elliptical perimeters, there arises his most important mathematical contribution, which is now called Landen's transformation. Many very important and seemingly unrelated guises of Landen's transformation exist in the literature. Thus, a fourth purpose of this article is to delineate several formulations of Landen's transformation as well as to provide a short biography of this undeservedly, rather obscure, mathematician.

For several years, Landen published almost exclusively in the *Ladies Diary*. This is, historically, the first regularly published periodical to contain a section devoted to the posing of mathematical problems and their solutions. Because an important feature of the MONTHLY has its roots in the *Ladies Diary*, it seems then dually appropriate in this paper to provide a brief description of the *Ladies Diary*.

2. Gauss and the arithmetic-geometric mean. As we previously alluded, the arithmetic-geometric mean was first set forth in a memoir of Lagrange **[30]** published in 1784–85. However, in a letter, dated April 16, 1816, to a friend, H. C. Schumacher, Gauss confided that he independently discovered the arithmetic-geometric mean in 1791 at the age of 14. At about the age of 22 or 23, Gauss wrote a long paper **[23]** describing his many discoveries on the arithmetic-geometric mean. However, this work, like many others by Gauss, was not published until after his death. Gauss's fundamental paper thus did not appear until 1866 when E. Schering, the editor of Gauss's complete works, published the paper as part of Gauss's *Nachlass*. Gauss obviously attached considerable importance to his findings on the arithmetic-geometric mean, for several of the entries in his diary, in particular, from the years 1799 to 1800, pertain to the arithmetic-geometric mean. Some of these entries are quite vague, and we may still not know everything that Gauss discovered about the arithmetic-geometric mean. (For an English translation of Gauss's diary together with commentary, see a paper by J. J. Gray **[24]**.)

By now, the reader is anxious to learn about the arithmetic-geometric mean and what the young Gauss discovered.

Let a and b denote positive numbers with $a > b$. Construct a sequence of arithmetic means and a sequence of geometric means as follows:

$$\begin{aligned} a_1 &= \frac{1}{2}(a+b), & b_1 &= \sqrt{ab}, \\ a_2 &= \frac{1}{2}(a_1+b_1), & b_2 &= \sqrt{a_1 b_1}, \\ &\vdots & &\vdots \\ a_{n+1} &= \frac{1}{2}(a_n+b_n), & b_{n+1} &= \sqrt{a_n b_n}, \\ &\vdots & &\vdots \end{aligned}$$

Gauss [23] gives four numerical examples, of which we reproduce one. Let $a = 1$ and $b = 0.8$. Then

$$\begin{aligned} a_1 &= 0.9, & b_1 &= 0.894427190999915878564, \\ a_2 &= 0.897213595499957939282, & b_2 &= 0.897209268732734, \\ a_3 &= 0.897211432116346, & b_3 &= 0.897211432113738, \\ a_4 &= 0.897211432115042, & b_4 &= 0.897211432115042. \end{aligned}$$

(Obviously, Gauss did not shirk from numerical calculations.) It appears from this example that $\{a_n\}$ and $\{b_n\}$ converge to the same limit, and that furthermore this convergence is very rapid. This we now demonstrate.

Observe that

$$\begin{gathered} b < b_1 < a_1 < a, \\ b < b_1 < b_2 < a_2 < a_1 < a, \\ b < b_1 < b_2 < b_3 < a_3 < a_2 < a_1 < a, \end{gathered}$$

etc. Thus, $\{b_n\}$ is increasing and bounded, and $\{a_n\}$ is decreasing and bounded. Each sequence therefore converges. Elementary algebraic manipulation now shows that

$$\frac{a_1 - b_1}{a - b} = \frac{a - b}{4(a_1 + b_1)} = \frac{a - b}{2(a + b) + 4b_1} < \frac{1}{2}.$$

Iterating this procedure, we deduce that

$$a_n - b_n < \left(\frac{1}{2}\right)^n (a - b), \qquad n \geqslant 1,$$

which tends to 0 as n tends to ∞. Thus, a_n and b_n converge to the same limit, which we denote by $M(a, b)$. By definition, *$M(a, b)$ is the arithmetic-geometric mean of a and b.*

To provide a more quantitative measure of the rapidity of convergence, first define

$$c_n = \sqrt{a_n^2 - b_n^2}, \qquad n \geqslant 0, \tag{1}$$

where $a_0 = a$ and $b_0 = b$. Observe that

$$c_{n+1} = \frac{1}{2}(a_n - b_n) = \frac{c_n^2}{4a_{n+1}} \leqslant \frac{c_n^2}{4M(a,b)}.$$

Thus, c_n tends to 0 quadratically, or the convergence is of the second order. More generally, suppose that $\{\alpha_n\}$ converges to L and assume that there exist constants $C > 0$ and $m \geqslant 1$ such that

$$|\alpha_{n+1} - L| \leqslant C|\alpha_n - L|^m, \qquad n \geqslant 1.$$

Then we say that the convergence is of the mth order.

Perhaps the most significant theorem in Gauss's paper [23] is the following representation for M for which we provide Gauss's ingenious proof.

THEOREM 1. *Let* $|x| < 1$, *and define*

$$K(x) = \int_0^{\pi/2} (1 - x^2\sin^2\varphi)^{-1/2}\, d\varphi. \tag{2}$$

Then

$$M(1+x, 1-x) = \frac{\pi}{2K(x)}.$$

The integral $K(x)$ is called the complete elliptic integral of the first kind. Observe that in the definition of $K(x)$, $\sin^2\varphi$ may be replaced by $\cos^2\varphi$.

Before proving Theorem 1, we give a reformulation of it. Define

$$I(a,b) = \int_0^{\pi/2} (a^2\cos^2\varphi + b^2\sin^2\varphi)^{-1/2}\, d\varphi. \tag{3}$$

It is easy to see that

$$I(a,b) = \frac{1}{a}K(x),$$

where

$$x = \frac{1}{a}\sqrt{a^2 - b^2}.$$

Since

$$M(a,b) = M(a_1, b_1) \quad \text{and} \quad M(ca, cb) = cM(a,b), \tag{4}$$

for any constant c, it follows that, with x as above,

$$M(1+x, 1-x) = \frac{1}{a}M(a,b).$$

The following reformulation of Theorem 1 is now immediate.

THEOREM 1′. *Let* $a > b > 0$. *Then*

$$M(a,b) = \frac{\pi}{2I(a,b)}.$$

Proof. Clearly, $M(1+x,1-x)$ is an even function of x. Gauss then *assumes* that

$$\frac{1}{M(1+x,1-x)}=\sum_{k=0}^{\infty}A_k x^{2k}. \tag{5}$$

Now make the substitution $x=2t/(1+t^2)$. From (4), it follows that

$$M(1+x,1-x)=\frac{1}{1+t^2}M\big((1+t)^2,(1-t)^2\big)=\frac{1}{1+t^2}M(1+t^2,1-t^2).$$

Substituting in (5), we find that

$$(1+t^2)\sum_{k=0}^{\infty}A_k t^{4k}=\sum_{k=0}^{\infty}A_k\left(\frac{2t}{1+t^2}\right)^{2k}.$$

Clearly, $A_0=1$. Expanding $(1+t^2)^{-2k-1}$, $k\geqslant 0$, in a binomial series and equating coefficients of like powers of t on both sides, we eventually find that

$$\begin{aligned}\frac{1}{M(1+x,1-x)}&=1+\left(\frac{1}{2}\right)^2x^2+\left(\frac{1\cdot 3}{2\cdot 4}\right)^2x^4+\left(\frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\right)^2x^6+\cdots\\&=\sum_{k=0}^{\infty}\frac{\left(\frac{1}{2}\right)_k^2}{(k!)^2}x^{2k}.\end{aligned} \tag{6}$$

Here we have introduced the notation

$$(\alpha)_k=\alpha(\alpha+1)(\alpha+2)\cdots(\alpha+k-1). \tag{7}$$

Complete details for the derivation of (6) may be found in Gauss's paper [**23**, pp. 367–369].

We now must identify the series in (6) with $K(x)$. Expanding the integrand of $K(x)$ in a binomial series and integrating termwise, we find that

$$\begin{aligned}K(x)&=\sum_{k=0}^{\infty}\frac{\left(\frac{1}{2}\right)_k}{k!}x^{2k}\int_0^{\pi/2}\sin^{2k}\varphi\,d\varphi\\&=\frac{\pi}{2}\sum_{k=0}^{\infty}\frac{\left(\frac{1}{2}\right)_k^2}{(k!)^2}x^{2k}.\end{aligned} \tag{8}$$

Combining (6) and (8), we complete the proof of Gauss's theorem.

Another short, elegant proof of Theorem 1 has been given by Newman [**39**] and is sketched by J. M. and P. B. Borwein [**10**].

For a very readable, excellent account of Gauss's many contributions to the arithmetic-geometric mean, see Cox's paper [**18**]. We shall continue the discussion of some of Gauss's discoveries in Section 5.

3. Landen and the Ladies Diary. We next sketch another proof of Theorem 1 (or Theorem 1′) which is essentially due to the eighteenth-century English mathematician John Landen.

Second proof. Although the basic idea is due to Landen, the iterative procedure that we shall describe is apparently due to Legendre [**33**, pp. 79–83] some years later.

For brevity, set

$$x_n = c_n/a_n, \qquad n \geqslant 0, \tag{9}$$

where c_n is defined by (1). In the complete elliptic integral of the first kind (2), make the substitution

$$\tan \varphi_1 = \frac{\sin(2\varphi)}{x_1 + \cos(2\varphi)}. \tag{10}$$

This is called Landen's transformation. After a considerable amount of work, we find that

$$K(x) = (1 + x_1)K(x_1).$$

Upon n iterations, we deduce that

$$K(x) = (1 + x_1)(1 + x_2) \cdots (1 + x_n)K(x_n). \tag{11}$$

Since, by (1) and (9), $1 + x_k = a_{k-1}/a_k$, $k \geqslant 1$, we see that (11) reduces to

$$K(x) = \frac{a}{a_n}K(x_n).$$

We now let n tend to ∞. Since a_n tends to $M(a, b)$ and x_n tends to 0, we conclude that

$$K(x) = \frac{a}{M(a, b)}K(0) = \frac{a\pi}{2M(a, b)}.$$

Landen's transformation (10) was introduced by him in a paper [**31**] published in 1771 and in more developed form in his most famous paper [**32**] published in 1775. There exist several versions of Landen's transformation. Often Landen's transformation is expressed as an equality between two differentials in the theory of elliptic functions [**17**], [**37**]. The importance of Landen's transformation is conveyed by Mittag-Leffler who, in his very perceptive survey [**37**, p. 291] on the theory of elliptic functions, remarks, "Euler's addition theorem and the transformation theorem of Landen and Lagrange were the two fundamental ideas of which the theory of elliptic functions was in possession when this theory was brought up for renewed consideration by Legendre in 1786."

In Section 4, we shall prove the following theorem, which is often called Landen's transformation for complete elliptic integrals of the first kind.

THEOREM 2. *If* $0 \leqslant x < 1$, *then*

$$K\left(\frac{2\sqrt{x}}{1 + x}\right) = (1 + x)K(x).$$

In fact, Theorem 2 is the special case $\alpha = \pi$, $\beta = \pi/2$ of the following more general formula. If $x \sin \alpha = \sin(2\beta - \alpha)$, then

$$(1 + x)\int_0^{\alpha}(1 - x^2\sin^2\varphi)^{-1/2}\,d\varphi = 2\int_0^{\beta}\left(1 - \frac{4x}{(1 + x)^2}\sin^2\varphi\right)^{-1/2} d\varphi,$$

which is known as Landen's transformation for incomplete elliptic integrals of the first kind.

To describe another form of Landen's transformation, we introduce Gauss's ordinary hypergeometric series

$$F(a, b; c; x) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k k!} x^k, \qquad |x| < 1, \tag{12}$$

where a, b, and c denote arbitrary complex numbers and $(\alpha)_k$ is defined by (7). Then

$$F\left(a, b; 2b; \frac{4x}{(1+x)^2}\right) = (1+x)^{2a} F\left(a, a-b+\frac{1}{2}; b+\frac{1}{2}; x^2\right) \tag{13}$$

is Landen's transformation for hypergeometric series. Theorems 1 and 2 imply the special case

$$F\left(\frac{1}{2}, \frac{1}{2}; 1; \frac{4x}{(1+x)^2}\right) = (1+x) F\left(\frac{1}{2}, \frac{1}{2}; 1; x^2\right).$$

Thus, a seemingly innocent "change of variable" (10) has many important ramifications. Indeed, Landen himself evidently never realized the importance of his idea.

Since Landen undoubtedly is not known to most readers, it seems appropriate here to give a brief biography. He was born in 1719. According to the Encyclopedia Britannica [**20**], "He lived a very retired life, and saw little or nothing of society; when he did mingle in it, his dogmatism and pugnacity caused him to be generally shunned." In 1762, he was appointed as the land-agent to the Earl Fitzwilliam, a post he held until two years before his death in 1790.

As a mathematician, Landen was primarily an analyst and geometer. Most of his important works were published in the latter part of his career. These include the aforementioned papers and *Mathematical Memoirs*, published in 1780 and 1789. For several years, Landen contributed many problems and solutions to the *Ladies Diary*. From 1743–1749, he posed a total of eleven problems and published thirteen solutions to problems. However, Leybourn [**34**] has disclosed that contributors to the *Ladies Diary* frequently employed aliases. In particular, Landen used the pseudonyms Sir Stately Stiff, Peter Walton, Waltoniensis, C. Bumpkin, and Peter Puzzlem, who, collectively, proposed ten problems and answered seventeen. Leybourn [**34**] has compiled in four volumes the problems and solutions from the *Ladies Diary* from 1704–1816. Especially valuable are his indices of subject classifications and contributors. (The problems and solutions from the years 1704–1760 had been previously collected by others in one volume in 1774 [**50**].)

First published in 1704, the annual *Ladies Diary* evidently was very popular in England with a yearly circulation of several thousand. The *Ladies Diary* is "designed principally for the amusement and instruction of the fair sex." It contains "new improvements in arts and sciences, and many entertaining particulars... for the use and diversion of the fair sex." The cover is graced by a poem dedicated to the reigning queen and which normally changed little from year to year. Our paper begins with the poem from 1776 paying eloquent homage to the beloved of King George III. Among other things, the *Ladies Diary* contains a "chronology of

remarkable events," birth dates of the royal family, enigmas, and answers to enigmas from the previous year. The enigmas as well as the answers were normally set to verse.

The largest portion of the *Ladies Diary* is devoted to the solutions of mathematical problems posed in the previous issue. Despite the name of the journal, very few contributors were women. Leybourn's **[34]** index lists a total of 913 contributors of which 32 were women. Because proposers and solvers did occasionally employ pen names such as Plus Minus, Mathematicus, Amicus, Archimedes, Diophantoides, and the aforementioned aliases for Landen, it is possible that the number of female contributors is slightly higher. In 1747, Landen gave a solution to a problem which was "designed to improve gunnery of which there are several things wanting." Does not this have a familiar ring today? Geometrical problems were popular, and rigor was lax at times. Here is an example from 1783. Let

$$a = \sum_{k=0}^{\infty} \frac{1}{\sqrt{2k+1}} \quad \text{and} \quad b = \sum_{k=1}^{\infty} \frac{1}{\sqrt{2k}}.$$

Show that $a/b = \sqrt{2} - 1$. In 1784, Joseph French provided the following "elegant" solution. We see that

$$\sqrt{2} \sum_{k=1}^{\infty} \frac{1}{\sqrt{2k}} = \sum_{k=1}^{\infty} \frac{1}{\sqrt{k}} = a + b.$$

Thus, $b(\sqrt{2} - 1) = a$, and the result follows.

Those readers wishing to learn more about Landen's work should consult Watson's very delightful article, "The Marquis and the Land-agent" **[52]**. Readers desiring more knowledge of the mathematical content of the *Ladies Diary* should definitely consult Leybourn's compendium **[34]**. (Only a few libraries in the U.S. possess copies of the *Ladies Diary*. The University of Illinois Library has a fairly complete collection, although there are several gaps prior to 1774. T. Perl **[43]** has written a detailed description of the *Ladies Diary* with an emphasis on the contributions by women and an analysis of both the positive and negative sociological factors on womens' mathematical education during the years of the *Diary*. For additional historical information about the *Ladies Diary* and other obscure English journals containing mathematics, see Archibald's paper **[2]**.)

4. Ivory and Landen's transformation. In 1796, J. Ivory **[25]** published a new formula for the perimeter of an ellipse. A very similar proof establishes Theorem 2, a version of Landen's transformation discussed in the previous section.

Before proving Theorem 2, we note that it implies a new version of Theorem 1.

THEOREM 1″. *If* $x > 0$, *then*

$$M(1+x, 1-x) = \frac{\pi(1+x)}{2K\left(\dfrac{2\sqrt{x}}{1+x}\right)}.$$

Theorem 1′ also follows from Theorem 1″; put $x = (a-b)/(a+b)$ and utilize (4).

Proof of Theorem 2. Using the definition (2) of K, employing the binomial series, and inverting the order of summation and integration below, we find that

$$\begin{aligned}
K\left(\frac{2\sqrt{x}}{1+x}\right) &= \frac{1}{2}\int_0^{\pi}\left(1-\frac{4x}{(1+x)^2}\sin^2\varphi\right)^{-1/2} d\varphi \\
&= \frac{1}{2}\int_0^{\pi}\left(1-\frac{2x}{(1+x)^2}(1-\cos(2\varphi))\right)^{-1/2} d\varphi \\
&= \frac{1}{2}(1+x)\int_0^{\pi}\left(1+x^2+2x\cos(2\varphi)\right)^{-1/2} d\varphi \\
&= \frac{1}{2}(1+x)\int_0^{\pi}\left(1+xe^{2i\varphi}\right)^{-1/2}\left(1+xe^{-2i\varphi}\right)^{-1/2} d\varphi \\
&= \frac{1}{2}(1+x)\sum_{m=0}^{\infty}\frac{\left(\frac{1}{2}\right)_m(-x)^m}{m!}\sum_{n=0}^{\infty}\frac{\left(\frac{1}{2}\right)_n(-x)^n}{n!}\int_0^{\pi}e^{2i(m-n)\varphi}\,d\varphi \\
&= \frac{\pi}{2}(1+x)\sum_{n=0}^{\infty}\frac{\left(\frac{1}{2}\right)_n^2 x^{2n}}{(n!)^2} \\
&= (1+x)K(x),
\end{aligned}$$

by (8). This concludes the proof.

Ivory's paper **[25]**, establishing an analogue of Theorem 2, possesses an unusual feature in that it begins with the "cover letter" that Ivory sent to the editor John Playfair when he submitted his paper! In this letter, Ivory informs Playfair about what led him to his discovery. Evidently then, the editor deemed it fair play to publish Ivory's letter as a preamble to his paper. The letter reads as follows.

Dear Sir,
Having, as you know, bestowed a good deal of time and attention on the study of that part of physical astronomy which relates to the mutual disturbances of the planets, I have, naturally, been led to consider the various methods of resolving the formula $(a^2 + b^2 - 2ab\cos\varphi)^n$ into infinite series of the form $A + B\cos\varphi + C\cos 2\varphi + \&c.$ In the course of these investigations, a series for the rectification of the ellipsis occurred to me, remarkable for its simplicity, as well as its rapid convergency. As I believe it to be new, I send it to you, inclosed, together with some remarks on the evolution of the formula just mentioned, which if you think proper, you may submit to the consideration of the Royal Society.

I am, Dear Sir,
Yours, & *c.*
James Ivory

5. Calculation of π. First, we define the complete elliptic integral of the second kind,

$$E(x) := \int_0^{\pi/2}\left(1-x^2\sin^2\varphi\right)^{1/2} d\varphi,$$

where $|x| < 1$. Two formulas relating the elliptic integrals $E(x)$ and $K(x)$ are the basis for one of the currently most efficient methods to calculate π. The first is due

to Legendre [33, p. 61]. We give below a simple proof that appears not to have been, heretofore, given.

THEOREM 3. *Let* $x' = \sqrt{1-x^2}$, *where* $0 < x < 1$. *Then*

$$K(x)E(x') + K(x')E(x) - K(x)K(x') = \frac{\pi}{2}. \tag{14}$$

Proof. Let $c = x^2$ and $c' = 1 - c$. A straightforward calculation gives

$$\begin{aligned}\frac{d}{dc}(E-K) &= -\frac{d}{dc}\int_0^{\pi/2}\frac{c\sin^2\varphi}{\left(1-c\sin^2\varphi\right)^{1/2}}\,d\varphi\\ &= \frac{E}{2c} - \frac{1}{2c}\int_0^{\pi/2}\frac{d\varphi}{\left(1-c\sin^2\varphi\right)^{3/2}}.\end{aligned}$$

Since

$$\frac{d}{d\varphi}\left(\frac{\sin\varphi\cos\varphi}{\left(1-c\sin^2\varphi\right)^{1/2}}\right) = \frac{1}{c}\left(1-c\sin^2\varphi\right)^{1/2} - \frac{c'}{c}\left(1-c\sin^2\varphi\right)^{-3/2},$$

we deduce that

$$\begin{aligned}\frac{d}{dc}(E-K) &= \frac{E}{2c} - \frac{E}{2cc'} + \frac{1}{2c'}\int_0^{\pi/2}\frac{d}{d\varphi}\left(\frac{\sin\varphi\cos\varphi}{\left(1-c\sin^2\varphi\right)^{1/2}}\right)d\varphi\\ &= \frac{E}{2c}\left(1-\frac{1}{c'}\right) = -\frac{E}{2c'}.\end{aligned} \tag{15}$$

For brevity, put $K' = K(c')$ and $E' = E(c')$. Since $c' = 1 - c$, it follows that

$$\frac{d}{dc}(E'-K') = \frac{E'}{2c}. \tag{16}$$

Lastly, easy calculations yield

$$\frac{dE}{dc} = \frac{E-K}{2c} \quad \text{and} \quad \frac{dE'}{dc} = -\frac{E'-K'}{2c'}. \tag{17}$$

If L denotes the left side of (14), we may write L in the form

$$L = EE' - (E-K)(E'-K').$$

Employing (15)–(17), we find that

$$\frac{dL}{dc} = \frac{(E-K)E'}{2c} - \frac{E(E'-K')}{2c'} + \frac{E(E'-K')}{2c'} - \frac{(E-K)E'}{2c} = 0.$$

Hence, L is a constant, and we will find its value by letting c approach 0.

First,

$$E - K = -c\int_0^{\pi/2}\frac{\sin^2\varphi}{\left(1-c\sin^2\varphi\right)^{1/2}}\,d\varphi = O(c)$$

as c tends to 0. Next,

$$K' = \int_0^{\pi/2} (1 - c'\sin^2\varphi)^{-1/2}\, d\varphi \leqslant \int_0^{\pi/2} (1 - c')^{-1/2}\, d\varphi$$
$$= O(c^{-1/2}),$$

as c tends to 0. Thus,

$$\lim_{c\to 0} L = \lim_{c\to 0} \{(E-K)K' + E'K\}$$
$$= \lim_{c\to 0} \left\{ O(c^{1/2}) + 1 \cdot \frac{\pi}{2} \right\} = \frac{\pi}{2},$$

and the proof is complete.

The second key formula, given in Theorem 4 below, can be proved via an iterative process involving Landen's transformation. We forego a proof here; a proof may be found, for example, in King's book [**29**, pp. 7, 8].

THEOREM 4. *Let, for* $a > b > 0$,

$$J(a, b) = \int_0^{\pi/2} (a^2\cos^2\varphi + b^2\sin^2\varphi)^{1/2}\, d\varphi, \tag{18}$$

and recall that c_n *is defined by* (1). *Then*

$$J(a, b) = \left(a^2 - \frac{1}{2} \sum_{n=0}^{\infty} 2^n c_n^2 \right) I(a, b),$$

where $I(a, b)$ *is defined by* (3).

Note that

$$J(a, b) = aE(x),$$

where $x = (1/a)\sqrt{a^2 - b^2}$.

Theorems 3 and 4 now lead to a formula for π which is highly suitable for computation.

THEOREM 5. *If* c_n *is defined by* (1), *then*

$$\pi = \frac{4M^2(1, 1/\sqrt{2})}{1 - \sum_{n=1}^{\infty} 2^{n+1} c_n^2}.$$

Proof. Letting $x = x' = 1/\sqrt{2}$ in Theorem 3, we find that

$$2K\left(\frac{1}{\sqrt{2}}\right)E\left(\frac{1}{\sqrt{2}}\right) - K^2\left(\frac{1}{\sqrt{2}}\right) = \frac{\pi}{2}. \tag{19}$$

Setting $a = 1$ and $b = 1/\sqrt{2}$ in Theorem 4, we see that

$$E\left(\frac{1}{\sqrt{2}}\right) = \left(1 - \frac{1}{2}\sum_{n=0}^{\infty} 2^n c_n^2\right) K\left(\frac{1}{\sqrt{2}}\right), \tag{20}$$

since $I(1, \sqrt{2}) = K(1/\sqrt{2})$ and $J(1, 1/\sqrt{2}) = E(1/\sqrt{2})$. Lastly, by Theorem 1′,

$$M(1, 1/\sqrt{2}) = \frac{\pi}{2K(1/\sqrt{2})}. \tag{21}$$

Substituting (20) into (19), employing (21), noting that $c_0^2 = 1/2$, and solving for π, we complete the proof.

According to King [**29**, pp. 8, 9, 12], an equivalent form of Theorem 5 was established by Gauss. Observe that in the proof of Theorem 5, we used only the special case $x = x' = 1/\sqrt{2}$ of Legendre's identity, Theorem 3. We would like to show now that this special case is equivalent to the formula

$$\int_0^1 \frac{dx}{\sqrt{1-x^4}} \int_0^1 \frac{x^2\,dx}{\sqrt{1-x^4}} = \frac{\pi}{4}, \tag{22}$$

first proved by Euler [**22**] in 1782. (Watson [**52**, p. 12] claimed that an equivalent formulation of (22) was earlier established by both Landen and Wallis, but we have been unable to verify this.) The former integral in (22) is one quarter of the arc length of the lemniscate given by $r^2 = \cos(2\varphi)$, $0 \leqslant \varphi \leqslant 2\pi$. The latter integral in (22) is intimately connected with the classical elastic curve. For a further elaboration of the connections of these two curves with the arithmetic-geometric mean, see Cox's paper [**18**].

In order to prove (22), make the substitution $x = \cos\varphi$. Then straightforward calculations yield

$$K\left(\frac{1}{\sqrt{2}}\right) = \sqrt{2}\int_0^1 \frac{dx}{\sqrt{1-x^4}}$$

and

$$2E\left(\frac{1}{\sqrt{2}}\right) - K\left(\frac{1}{\sqrt{2}}\right) = \sqrt{2}\int_0^1 \frac{x^2\,dx}{\sqrt{1-x^4}}.$$

It is now easy to see that Legendre's relation Theorem 3 in the case $x = x' = 1/\sqrt{2}$ implies (22).

In 1976, Salamin [**47**] rederived the forgotten Theorem 5, from which he established a rapidly convergent algorithm for the computation of π. Recall that in Section 2 we demonstrated how rapidly the arithmetic-geometric mean converges and thus how fast c_n tends to 0. Tamura and Kanada have used this algorithm to compute π to 2^{24} (over 16 million) decimal places. An announcement about their calculation of π to 2^{23} decimal places was made in *Scientific American* [**48**]. Their paper [**27**] describes their calculation to 10,013,395 decimal places. More recently, D. H. Bailey [**3**] has used a quartically convergent algorithm to calculate π to 29,360,000 digits.

Newman [**40**] has obtained a quadratic algorithm for the computation of π that is somewhat simpler than Salamin's. His proof is quite elementary and avoids Legendre's identity. It should be remarked that Newman's estimates of some integrals are not quite correct. However, the final result (middle of p. 209) is correct.

In 1977, Brent [**14**] observed that the arithmetic-geometric mean could be implemented to calculate elementary functions as well. Let us briefly indicate how

to calculate $\log 2$. From Whittaker and Watson's text [**53**, p. 522], as x tends to $1-$,

$$K(x) \sim \log\frac{4}{\sqrt{1-x^2}}.$$

From Theorem 1′,

$$K(x) = \frac{\pi}{2M\left(1, \sqrt{1-x^2}\right)},$$

and so

$$\log\frac{4}{\sqrt{1-x^2}} \approx \frac{\pi}{2M\left(1, \sqrt{1-x^2}\right)}.$$

Taking $\sqrt{1-x^2} = 4 \cdot 2^{-n}$, we find that

$$\log 2 \approx \frac{\pi}{2nM\left(1, 2^{2-n}\right)},$$

for large n.

Further improvements in both the calculation of π and elementary functions have been made by J. M. and P. B. Borwein [**8**], [**9**], [**10**], [**11**], [**12**], [**13**]. In particular, in [**8**], [**11**], and [**12**], they have utilized elliptic integrals and modular equations to obtain algorithms of higher order convergence to approximate π. The survey article [**10**] by the Borwein brothers is to be especially recommended. Carlson [**16**] has written an earlier survey on algorithms dependent on the arithmetic-geometric mean and variants thereof.

Postscript to π. The challenge of approximating and calculating π has been with us for over 4000 years. By 1844, π was known to 200 decimal places. This stupendous feat was accomplished by a calculating prodigy named Johann Dase in less than two months. On Gauss's recommendation, the Hamburg Academy of Sciences hired Dase to compute the factors of all integers between 7,000,000 and 10,000,000. Thus, our ideas have come to a full circle. As Beckmann [**5**, p. 104] remarks, "It would thus appear that Carl Friedrich Gauss, who holds so many firsts in all branches of mathematics, was also the first to introduce payment for computer time." The computer time now for 29 million digits (28 hours) is considerably less than the computer time for 200 digits by Gauss's computer, Dase.

6. Approximations for the perimeter L of an ellipse. If an ellipse is given by the parametric equations $x = a\cos\varphi$ and $y = b\sin\varphi$, $0 \leqslant \varphi \leqslant 2\pi$, then from elementary calculus,

$$\begin{aligned} L = L(a,b) &= \int_0^{2\pi}\left(a^2\sin^2\varphi + b^2\cos^2\varphi\right)^{1/2} d\varphi \\ &= 4J(b,a), \end{aligned} \tag{23}$$

where $J(b, a)$ is defined by (18). Thus, we see immediately from Theorems 1′ and 4 that elliptical perimeters and arithmetic-geometric means are inextricably intertwined. Ivory's letter and our concomitant comments also unmistakenly pointed to this union.

Before discussing approximations for L, we offer two exact formulas. The former is due to MacLaurin [36] in 1742, and the latter was initially found by Ivory [25] in 1796, although it is implicit in the earlier work of Landen.

THEOREM 6. *Let* $x = a\cos\varphi$ *and* $y = b\sin\varphi$, $0 \leqslant \varphi \leqslant 2\pi$. *Let* $e = (1/a)\sqrt{a^2 - b^2}$, *the eccentricity of the ellipse. Then if F is defined by* (12),

$$L(a,b) = 2\pi a F\left(\frac{1}{2}, -\frac{1}{2}; 1; e^2\right) \tag{24}$$

$$= \pi(a+b) F\left(-\frac{1}{2}, -\frac{1}{2}; 1; \lambda^2\right), \tag{25}$$

where

$$\lambda = \frac{a-b}{a+b}.$$

Proof. The proofs are very similar to those in Sections 2 and 4. First, using (23), expanding the integrand in a binomial series, and integrating termwise, we deduce that

$$\begin{aligned} L(a,b) &= 4a\int_0^{\pi/2}\left(1 - e^2\cos^2\varphi\right)^{1/2} d\varphi \\ &= 4a\sum_{n=0}^{\infty}\frac{\left(-\frac{1}{2}\right)_n}{n!}e^{2n}\int_0^{\pi/2}\cos^{2n}\varphi\, d\varphi \qquad (26) \\ &= 2\pi a F\left(\frac{1}{2}, -\frac{1}{2}; 1; e^2\right). \end{aligned}$$

Thus, (24) is established.

We indicate two proofs of (25). First, in Landen's transformation (13) of hypergeometric series, set $a = -1/2$, $b = 1/2$, and $x = \lambda$. We immediately find that

$$F\left(-\frac{1}{2}, \frac{1}{2}; 1; e^2\right) = \frac{a+b}{2a} F\left(-\frac{1}{2}, -\frac{1}{2}; 1; \lambda^2\right).$$

By this formula and (24), formula (25) is demonstrated.

The second proof that we mention is that of Ivory [25]. Using (26), proceed exactly in the same fashion as in the proof of Theorem 2 in Section 4.

In fact, there exists a third early formula for $L(a,b)$. In 1773, Euler [21] proved that

$$L(a,b) = \pi\sqrt{2(a^2+b^2)}\, F\left(-\frac{1}{4}, \frac{1}{4}; 1; \left(\frac{a^2-b^2}{a^2+b^2}\right)^2\right).$$

Although Euler proceeded differently, we mention that his formula may be derived from MacLaurin's via a certain quadratic transformation for hypergeometric series that is different from Landen's. Euler's formula also trivially leads to an approximation for $L(a,b)$ given in our table below.

The problem of determining $L(a, b)$ is not as venerable as that for determining π. However, some have argued (not very convincingly) that the problem goes back to the time of King Solomon, who hired a craftsman Huram to make a tank. According to 1 Kings 7 : 23, "Huram made a round tank of bronze 5 cubits deep, 10 cubits in diameter, and 30 cubits in circumference." The implication is clear that the ancient Hebrews regarded π as being equal to 3. It has been suggested, perhaps by someone who believes that "God makes no mistakes," that "round" and "depth" are to be interpreted loosely, and that the tank really was elliptical in shape, with the major axis being 10 cubits and the minor axis being about 9.53 cubits in length.

As might be expected. the primary impetus in finding methods for calculating elliptical perimeters arises from astronomy. In 1609, Kepler **[28]** offered perhaps the first legitimate approximations

$$L \approx \pi(a+b) \quad \text{and} \quad L \approx 2\pi\sqrt{ab},$$

although, as we shall see, his arguments were not very rigorous and $2\pi\sqrt{ab}$ was intended to be only a *lower bound* for L. Kepler **[28**, p. 307] first remarks that the ellipse with semiaxes a and b and the circle with radius $\sqrt{ab}$ have the same areas. Since the circle has the smaller circumference,

$$L \geqslant 2\pi\sqrt{ab}.$$

He **[28**, p. 368] furthermore remarks that $(1/2)(a+b) \geqslant \sqrt{ab}$, and so concludes that

$$L \approx 2\pi\frac{1}{2}(a+b).$$

Kepler appears to be using the dubious principle that quantities larger than the same number must be about equal.

Approximations of several types, depending upon the relative sizes of a and b, exist in the literature. In this section, we concentrate on estimates that are best for a close to b. Thus, we shall write all of our approximations in terms of $\lambda = (a-b)/(a+b)$ and compare them with the expansion (25). For example, Kepler's second approximation can be written in the form

$$L \approx \pi(a+b)(1-\lambda^2)^{1/2}.$$

We now show how the formula

$$L(a,b) = 4J(a,b) = \frac{2\pi}{M(a,b)}\left(a^2 - \frac{1}{2}\sum_{n=0}^{\infty} 2^n c_n^2\right), \tag{27}$$

arising from Theorems 1′ and 4, can be used to find approximations to the perimeter of an ellipse. Replacing $M(a, b)$ by a_2 and neglecting the terms with $n \geqslant 2$, we find that

$$L(a,b) \approx \frac{2\pi}{a_2}\left(a^2 - \frac{c_0^2}{2} - c_1^2\right) = \frac{2\pi a_1^2}{a_2} = 2\pi\left(\frac{a+b}{\sqrt{a}+\sqrt{b}}\right)^2.$$

This formula was first obtained by Ekwall **[19]** in 1973 as a consequence of a formula by Sipos from 1792 **[54]**.

If we replace $M(a, b)$ by a_3 in (27) and neglect all terms with $n \geqslant 3$, we find, after some calculation, that

$$L(a,b) \approx 2\pi \frac{2(a+b)^2 - (\sqrt{a} - \sqrt{b})^4}{(\sqrt{a} + \sqrt{b})^2 + 2\sqrt{2}\sqrt{a+b}\sqrt[4]{ab}}.$$

This formula is complicated enough to dissuade us from calculating further approximations by this method.

We now provide a table of approximations for $L(a, b)$ that have been given in the literature. At the left, we list the discoverer (or source) and year of discovery (if known). The approximation $A(\lambda)$ for $L(a, b)/\pi(a + b)$ is given in the second column in two forms. In the last column, the first nonzero term in the power series for

$$A(\lambda) - \frac{L(a,b)}{\pi(a+b)} = A(\lambda) - F\left(-\frac{1}{2}, -\frac{1}{2}; 1; \lambda^2\right)$$

is offered so that the accuracy of the approximating formula can be discerned. For convenience, we note that

$$F\left(-\frac{1}{2}, -\frac{1}{2}; 1; \lambda^2\right) = 1 + \frac{1}{4}\lambda^2 + \frac{1}{4^3}\lambda^4 + \frac{1}{4^4}\lambda^6 + \frac{25}{4^7}\lambda^8 + \frac{49}{4^8}\lambda^{10} + \cdots.$$

Kepler [**28**], 1609	$\frac{2\sqrt{ab}}{a+b} = (1-\lambda^2)^{1/2}$	$-\frac{3}{4}\lambda^2$
Euler [**21**], 1773	$\frac{\sqrt{2(a^2+b^2)}}{a+b} = (1+\lambda^2)^{1/2}$	$\frac{1}{4}\lambda^2$
Sipos [**54**], 1792 Ekwall [**19**], 1973	$\frac{2(a+b)}{(\sqrt{a}+\sqrt{b})^2} = \frac{2}{1+\sqrt{1-\lambda^2}}$	$\frac{7}{64}\lambda^4$
Peano [**42**], 1889	$\frac{3}{2} - \frac{\sqrt{ab}}{a+b} = \frac{3}{2} - \frac{1}{2}(1-\lambda^2)^{1/2}$	$\frac{3}{64}\lambda^4$
Muir [**38**], 1883	$\frac{2}{a+b}\left(\frac{a^{3/2}+b^{3/2}}{2}\right)^{2/3}$ $= \frac{1}{2^{2/3}}\{(1+\lambda)^{3/2}+(1-\lambda)^{3/2}\}^{2/3}$	$-\frac{1}{64}\lambda^4$
Lindner [**35**, p. 439], 1904–1920 Nyvoll [**41**], 1978	$\left\{1+\frac{1}{8}\left(\frac{a-b}{a+b}\right)^2\right\}^2$ $= \left(1+\frac{1}{8}\lambda^2\right)^2$	$-\frac{1}{2^8}\lambda^6$
Selmer [**49**], 1975	$1 + \frac{4(a-b)^2}{(5a+3b)(3a+5b)}$ $= 1 + \frac{1}{4}\lambda^2 \frac{1}{1-\frac{1}{16}\lambda^2}$	$-\frac{3}{2^{10}}\lambda^6$

Ramanujan [44], [45], 1914 Fergestad [49], 1951	$\dfrac{3-\sqrt{(a+3b)(3a+b)}}{a+b}$ $=3-\sqrt{4-\lambda^2}$	$-\dfrac{1}{2^9}\lambda^6$
Almkvist [1], 1978	$2\dfrac{2(a+b)^2-(\sqrt{a}-\sqrt{b})^4}{(a+b)\left\{(\sqrt{a}+\sqrt{b})^2+2\sqrt{2}\sqrt{a+b}\sqrt[4]{ab}\right\}}$ $=2\dfrac{\left(1+\sqrt{1-\lambda^2}\right)^2+\lambda^2\sqrt{1-\lambda^2}}{\left(1+\sqrt{1-\lambda^2}\right)\left(1+\sqrt[4]{1-\lambda^2}\right)^2}$	$\dfrac{15}{2^{14}}\lambda^8$
Bronshtein and Semendyayev [15], 1964 Selmer [49], 1975	$\dfrac{1}{16}\dfrac{64(a+b)^4-3(a-b)^4}{(a+b)^2(3a+b)(a+3b)}$ $=\dfrac{64-3\lambda^4}{64-16\lambda^2}$	$-\dfrac{9}{2^{14}}\lambda^8$
Selmer [49], 1975	$\dfrac{1}{8}\left\{12+\left(\dfrac{a-b}{a+b}\right)^2-\dfrac{2\sqrt{2(a^2+6ab+b^2)}}{a+b}\right\}$ $=\dfrac{3}{2}+\dfrac{1}{8}\lambda^2-\dfrac{1}{2}\sqrt{1-\dfrac{1}{2}\lambda^2}$	$-\dfrac{5}{2^{14}}\lambda^8$
Jacobsen and Waadeland [26], 1985	$\dfrac{256-48\lambda^2-21\lambda^4}{256-112\lambda^2+3\lambda^4}$	$-\dfrac{33}{2^{18}}\lambda^{10}$
Ramanujan [44], [45], 1914	$1+\dfrac{3\lambda^2}{10+\sqrt{4-3\lambda^2}}$	$-\dfrac{3}{2^{17}}\lambda^{10}$

The two approximations by India's great mathematician, S. Ramanujan, were first stated by him in his notebooks [46, p. 217], and then later at the end of his paper [44], [45, p. 39], where he says that they were discovered empirically. Ramanujan [44], [45] also provides error approximations, but they are in a form different from that given here. Since

$$\lambda=\frac{a-b}{a+b}=\frac{1-\sqrt{1-e^2}}{1+\sqrt{1-e^2}}\approx\frac{e^2}{4},$$

we find that, for the first approximation,

$$\pi(a+b)\frac{\lambda^6}{2^9}\approx\pi a\left(1+\sqrt{1-e^2}\right)\frac{\left(e^2/4\right)^6}{2^9}<2\pi a\frac{e^{12}}{2^{21}}=\pi a\frac{e^{12}}{2^{20}},$$

which is the approximate error given by Ramanujan. Similarly, for the second approximation, Ramanujan states that the error is approximately equal to

$$3\pi a\frac{e^{20}}{2^{36}},$$

which is in agreement with our claim. The exactness of Ramanujan's second formula for eccentricities that are not too large is very good. For example, for the orbit of

Mercury ($e = 0.206$), the absolute error is about 1.5×10^{-13} meters. Note that if we set $b = 0$ in Ramanujan's second formula, we find that $\pi \approx 22/7$.

Fergestad [49] rediscovered Ramanujan's first formula several years later.

Despite Ramanujan's remark on the discovery of these two formulas, Jacobsen and Waadeland [26] have offered a very plausible explanation of Ramanujan's approximations. We confine our attention to the latter approximation, since the arguments are similar. Write

$$F\left(-\frac{1}{2}, -\frac{1}{2}; 1; \lambda^2\right) = 1 + \frac{\lambda^2}{4(1+w)}. \tag{28}$$

Then it can be shown that w has the continued fraction expansion

$$w = \frac{1}{3}\left\{\frac{-\frac{3}{16}\lambda^2}{1} + \frac{-\frac{3}{16}\lambda^2}{1} + \frac{-\frac{3}{16}\lambda^2}{1} + \frac{-\frac{11}{48}\lambda^2}{1} + \cdots\right\}.$$

If each numerator above is replaced by $-3\lambda^2/16$, then we obtain the approximation

$$w \approx \frac{1}{12}\left(-2 + \sqrt{4 - 3\lambda^2}\right).$$

Substituting this approximation in (28) and then using (25), we are immediately led to the estimate

$$\frac{L(a,b)}{\pi(a+b)} \approx 1 + \frac{3\lambda^2}{10 + \sqrt{4 - 3\lambda^2}}.$$

Since Ramanujan's facility in representing analytic functions by continued fractions is unmatched in mathematical history, it seems likely that Ramanujan discovered his approximations in this manner.

In the next section, we examine some approximations for $L(a, b)$ of a different type given by Ramanujan in his notebooks [46].

7. Further approximations given by Ramanujan. In his notebooks [46], Ramanujan offers some very unusual formulas, expressed in sexagesimal notation, for $L(a, b)$. The first is related to his approximation $3 - \sqrt{4 - \lambda^2}$ given in Section 6.

THEOREM 7. *Put*

$$L(a,b) = \pi(a+b)\left(1 + 4\sin^2\frac{1}{2}\theta\right), \quad 0 \leqslant \theta \leqslant \pi/4, \tag{29}$$

and

$$\sin\theta = \lambda\sin\alpha, \qquad \lambda = \frac{a-b}{a+b}. \tag{30}$$

Then, when the eccentricity $e = 1$, $\alpha = 30°18'6''$, *and as* e *tends to* 0, α *tends monotonically to* $30°$.

It is not clear how Ramanujan was led to this very unusual theorem. The variance of α over such a small interval is curious.

Proof. We shall prove Theorem 7 except for the conclusion about monotonicity. However, we shall show that $\alpha \geqslant \pi/6$ always.

For brevity, we write (25) in the form

$$\frac{L(a,b)}{\pi(a+b)}=\sum_{n=0}^{\infty}\alpha_n\lambda^{2n},\qquad |\lambda|<1. \tag{31}$$

It then follows from (29) and (30) that

$$3-2\sqrt{1-\lambda^2\sin^2\alpha}=1+4\sin^2\frac{1}{2}\theta=\sum_{n=0}^{\infty}\alpha_n\lambda^{2n},\qquad |\lambda|<1. \tag{32}$$

Next set

$$3-\sqrt{4-\lambda^2}=\sum_{n=0}^{\infty}\beta_n\lambda^{2n},\qquad |\lambda|<2. \tag{33}$$

As implied in Section 6, $\alpha_n=\beta_n$, $n=0,1,2$. We shall further show that, for $n\geqslant 3$,

$$\beta_n\leqslant\alpha_n/2^{n-2}. \tag{34}$$

From the definitions (31) and (33), respectively, short calculations show that, for $n\geqslant 1$,

$$\frac{\alpha_{n+1}}{\alpha_n}=\frac{(2n-1)^2}{(2n+2)^2}\quad\text{and}\quad\frac{\beta_{n+1}}{\beta_n}=\frac{2n-1}{8(n+1)}.$$

Thus,

$$\frac{\beta_{n+1}}{\beta_n}\bigg/\frac{\alpha_{n+1}}{\alpha_n}=\frac{n+1}{2(2n-1)}\leqslant\frac{1}{2},$$

if $n\geqslant 2$. Proceeding by induction, we deduce that

$$\frac{\beta_{n+1}}{\alpha_{n+1}}\leqslant\frac{1}{2}\frac{\beta_n}{\alpha_n}\leqslant\frac{1}{2^{n-1}},$$

for $n\geqslant 2$, and the proof of (34) is complete.

From (32) and (34), it follows that

$$3-\sqrt{4-\lambda^2}\leqslant 3-2\sqrt{1-\lambda^2\sin^2\alpha}.$$

Solving this inequality, we find that $\sin^2\alpha\geqslant 1/4$, or $\alpha\geqslant\pi/6$.

Second, we calculate α when $e=1$. Thus, $\lambda=1$ and $\theta=\alpha$. Therefore, from (25) and (32),

$$1+4\sin^2\frac{1}{2}\alpha=F\left(-\frac{1}{2},-\frac{1}{2};1;1\right)=\frac{4}{\pi}. \tag{35}$$

This evaluation follows from a general theorem of Gauss on the evaluation of hypergeometric series at the argument 1 [**4**, p. 2]. Moreover, this particular series is found in Gauss's diary under the date June, 1798 [**24**]. Thus,

$$\sin^2\frac{1}{2}\alpha=\frac{1}{\pi}-\frac{1}{4}=0.0683098861.$$

It follows that $\alpha=30°18'6''$.

Third, we calculate α when $e = 0$. From (30) and (32),

$$\lim_{\lambda\to 0} \sin^2\alpha = \lim_{\lambda\to 0} \frac{\sin^2\theta}{\lambda^2} = \lim_{\lambda\to 0} \frac{4\sin^2\frac{1}{2}\theta}{\lambda^2}$$
$$= \lim_{\lambda\to 0} \lambda^{-2} \sum_{n=1}^{\infty} \alpha_n \lambda^{2n} = \alpha_1 = \frac{1}{4}.$$

Thus, α tends to $\pi/6$ as e tends to 0.

Ramanujan [**46**, p. 224] offers another theorem, which we do not state, like Theorem 7 but which appears to be motivated by his second approximation for $L(a, b)$.

Ramanujan [**46**, p. 224] states two additional formulas each of which combines two approximations, one for e near 0 and the other for e close to 1. Again, we give just one of the pair. A complete proof of Theorem 8 below would be too lengthy for this paper, and so we shall just sketch the main ideas of the proof. Complete details may be found in [7].

THEOREM 8. *Set*

$$L(a,b) = \pi(a+b)\frac{\tan\theta}{\theta}, \qquad 0 \leqslant \theta < \pi/2, \tag{36}$$

and

$$\tan\theta = \lambda\cos\alpha, \qquad \lambda = \frac{a-b}{a+b}. \tag{37}$$

Then as e increases from 0 *to* 1, α *decreases from* $\pi/6$ *to* 0. *Furthermore*, α *is approximately given by*

$$\frac{2\sqrt{ab}}{a+b}\left\{30^\circ + 6^\circ 18' 49'' \frac{(\sqrt{a}-\sqrt{b})^2}{a+b} - 1^\circ 10' 55'' \left(\frac{a-b}{a+b}\right)^2\right\}. \tag{38}$$

Proof. If $e = 0$, then $\lambda = 0$ and $\theta = 0$. The argument is very similar to that in the proof of Theorem 7, and we find that

$$\lim_{\lambda\to 0} \cos^2\alpha = 3\alpha_1 = 3/4.$$

Thus, $\alpha = \pi/6$ when $\lambda = 0 = e$.

We next determine α when $e = 1$. Thus, $\lambda = 1$ and $\tan\theta = \cos\alpha$ by (37). From (25), (36), and (35),

$$\left.\frac{\tan\theta}{\theta}\right|_{\lambda=1} = F\left(-\frac{1}{2}, -\frac{1}{2}; 1; 1\right) = \frac{4}{\pi}.$$

Thus, $\theta = \pi/4$ and $\alpha = 0$.

It appears to be extremely difficult to show that as λ goes from 0 to 1, α *monotonically* decreases from $\pi/6$ to 0. It can be shown [7], however, that $0 \leqslant \alpha \leqslant \pi/6$, always. A proof depends upon a continued fraction for $\tan^{-1}x$.

The proof of (38) is very difficult, and we provide only a brief sketch. We observe (again) that

$$\sqrt{1-\lambda^2} = \frac{2\sqrt{ab}}{a+b},$$

and so

$$\sqrt{1-\lambda^2} - (1-\lambda^2) = \frac{2\sqrt{ab}}{(a+b)^2}(\sqrt{a}-\sqrt{b})^2.$$

Thus, Ramanujan is attempting to find an approximation to α of the form

$$\sqrt{1-\lambda^2}\left(A + B\{1-\sqrt{1-\lambda^2}\} + C\lambda^2\right), \tag{39}$$

which will be a good approximation both when λ is close to 0 and when λ is near 1. Our task is then to determine A, B, and C.

With a considerable amount of effort, it can be shown that [7]

$$\alpha = \frac{\pi}{6} - \frac{21\sqrt{3}}{160}\lambda^2 + O(\lambda^4) \tag{40}$$

in a neighborhood of $\lambda = 0$. The proper expansion near $\lambda = 1$ is even more difficult to obtain because $F(-\frac{1}{2}, -\frac{1}{2}; 1; \lambda^2)$ is not analytic at $\lambda = 1$. However, there does exist an asymptotic expansion for $F(-\frac{1}{2}, -\frac{1}{2}; 1; \lambda^2)$ as λ tends to $1-$, and employing this, we can show that [7]

$$\alpha = \sqrt{\frac{4-\pi}{2\pi-4}}\sqrt{1-\lambda^2} + o\left(\sqrt{1-\lambda^2}\right), \tag{41}$$

as λ tends to $1-$.

Having omitted the hard analysis, we now determine A, B, and C from (40) and (41) with little difficulty. When λ tends to 0, (39) tends to A. Thus, $A = \pi/6$, by (40). Next, examine $(\alpha - \pi/6)/\lambda^2$ as λ tends to 0. From (39) and (40), we find that

$$-\frac{\pi}{12} + \frac{1}{2}B + C = -\frac{21\sqrt{3}}{160}.$$

Now check $\alpha/\sqrt{1-\lambda^2}$ as λ tends to $1-$. From (39) and (41), we see that

$$\frac{\pi}{6} + B + C = \sqrt{\frac{4-\pi}{2\pi-4}}.$$

Simultaneously solving these last two equalities, we conclude that

$$B = 2\sqrt{\frac{4-\pi}{2\pi-4}} + \frac{21\sqrt{3}}{80} - \frac{\pi}{2} = 0.1101935$$

and

$$C = \frac{\pi}{3} - \sqrt{\frac{4-\pi}{2\pi-4}} - \frac{21\sqrt{3}}{80} = -0.0206291.$$

Converting A, B, and C to the sexagesimal system and substituting in (39), we complete the proof.

Although Ramanujan is well known for his approximations and asymptotic formulas in number theory, he has not been adequately recognized for his deep contributions to approximations and asymptotic series in analysis, because the vast majority of his results in the latter field have been hidden in his notebooks. These notebooks were begun in about 1903, when he was 15 or 16, and are a compilation of his mathematical discoveries without proofs. The last entries were made in 1914, when he sailed to England at the urging of G. H. Hardy. Although the editing of Ramanujan's notebooks was strongly advocated by Hardy and others immediately after Ramanujan's death in 1920, it is only recently that this has come to fruition **[6]**.

We have not attempted to give complete proofs of some of the theorems that we have described, but we hope that the principal ideas have been made clear. We have seen that a chain of related ideas stretches back over a period exceeding two centuries and provides impetus to contemporary mathematics. Ideas and topics that appear disparate are found to have common roots and merge together. For further elaboration of these ideas, readers should consult the works cited, especially Cox's paper **[18]**, the papers and book of J. M. and P. B. Borwein **[8]**–**[13]**, a paper by Almkvist **[1]** written in Swedish, and Berndt's forthcoming book **[7]**.

We are most grateful to Birger Ekwall for providing some very useful references.

REFERENCES

1. G. Almkvist, Aritmetisk-geometriska medelvärdet och ellipsens båglängd, *Nordisk Mat. Tidskr.*, 25–26 (1978) 121–130.
2. R. C. Archibald, Notes on some minor English mathematical serials, *Math. Gaz.*, 14 (1929) 379–400.
3. D. H. Bailey, The computation of π to 29,360,000 decimal digits using Borweins' quartically convergent algorithm, to appear.
4. W. N. Bailey, Generalized Hypergeometric Series, Stechert-Hafner, New York, 1964.
5. P. Beckmann, A history of π, Golem Press, Boulder, 1970.
6. B. C. Berndt, Ramanujan's Notebooks, Part I, Springer-Verlag, New York, 1985.
7. B. C. Berndt, Ramanujan's Notebooks, Part III, Springer-Verlag, New York, to appear.
8. J. Borwein, Some modular identities of Ramanujan useful in approximating π, *Proc. Amer. Math. Soc.*, 95 (1985) 365–371.
9. J. M. and P. B. Borwein, A very rapidly convergent product expansion for π, *BIT*, 23 (1983) 538–540.
10. J. M. and P. B. Borwein, The arithmetic-geometric mean and fast computation of elementary functions, *SIAM Review*, 26 (1984) 351–366.
11. J. M. and P. B. Borwein, Cubic and higher order algorithms for π, *Canad. Math. Bull.*, 27 (1984) 436–443.
12. J. M. and P. B. Borwein, Elliptic integrals and approximations to π, to appear.
13. J. M. and P. B. Borwein, Pi and the AGM—A Study in Analytic Number Theory and Computational Complexity, John Wiley, New York, 1987.
14. R. P. Brent, Fast multiple-precision evaluation of elementary functions, *J. Assoc. Comput. Mach.*, 23 (1976) 242–251.
15. I. N. Bronshtein and K. A. Semendyayev, A Guide-Book to Mathematics for Technologists and Engineers, trans. by J. Jaworowski and M. N. Bleicher, Pergamon Press, Macmillan, New York, 1964.
16. B. C. Carlson, Algorithms involving arithmetic and geometric means, *Amer. Math. Monthly*, 78 (1971) 496–505.
17. A. Cayley, An Elementary Treatise on Elliptic Functions, second ed., Dover, New York, 1961.

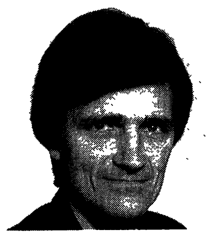
18. D. A. Cox, The arithmetic-geometric mean of Gauss, *L'Enseign. Math.*, 30 (1984) 275–330.
19. B. Ekwall, Approximationsformler för ellipsens omkrets, unpublished manuscript, 1973.
20. Encyclopedia Britannica, ninth ed., vol. 14, John Landen, Adam and Charles Black, Edinburgh, 1882, p. 271.
21. L. Euler, Nova series infinita maxime convergens perimetrum ellipsis exprimens, *Novi Comm. Acad. Sci. Petropolitanae* 18 (1773), 71–84; Opera Omnia, t.20, B. G. Teubner, Leipzig, 1912, pp. 357–370.
22. L. Euler, De miris proprietatibus curvae elasticae sub aequatione $y = \int xx\,dx/\sqrt{1-x^4}$ contentae, *Acta Acad. Sci. Petropolitanae* 1782: II (1786), 34–61; Opera Omnia, t.21, B. G. Teubner, Leipzig, 1913, pp. 91–118.
23. C. F. Gauss, Nachlass. Arithmetisch geometrisches Mittel, Werke, Bd. 3, Königlichen Gesell. Wiss., Göttingen, 1876, pp. 361–403.
24. J. J. Gray, A commentary on Gauss's mathematical diary, 1796–1814, with an English translation, *Expos. Math.*, 2 (1984) 97–130.
25. J. Ivory, A new series for the rectification of the ellipsis; together with some observations on the evolution of the formula $(a^2 + b^2 - 2ab\cos\varphi)^n$, *Trans. Royal Soc. Edinburgh*, 4 (1796) 177–190.
26. L. Jacobsen and H. Waadeland, Glimt fra analytisk teori for kjedebrøker, Del II, *Nordisk Mat. Tidskr.*, 33 (1985) 168–175.
27. Y. Kanada, Y. Tamura, S. Yoshino and Y. Ushiro, Calculation of π to 10,013,395 decimal places based on the Gauss-Legendre algorithm and Gauss arctangent relation, *Math. Comp.*, to appear.
28. J. Kepler, Opera Omnia, vol. 3, Astronomia Nova, Heyder & Zimmer, Frankfurt, 1860.
29. L. V. King, On the Direct Numerical Calculation of Elliptic Functions and Integrals, Cambridge University Press, Cambridge, 1924.
30. J.-L. Lagrange, Sur une novelle méthode de calcul intégral pour les différentielles affectées d'un radical carré sous lequel la variable ne passe pas le quatrième degré, *Mem. l'Acad. Roy. Sci. Turin* 2 (1784–85); Oeuvres, t.2, Gauthier-Villars, Paris, 1868, pp. 251–312.
31. J. Landen, A disquisition concerning certain fluents, which are assignable by the arcs of the conic sections; wherein are investigated some new and useful theorems for computing such fluents, *Philos. Trans. Royal Soc. London*, 61 (1771) 298–309.
32. J. Landen, An investigation of a general theorem for finding the length of any arc of any conic hyperbola, by means of two elliptic arcs, with some other new and useful theorems deduced therefrom, *Philos. Trans. Royal Soc. London*, 65 (1775) 283–289.
33. A. M. Legendre, Traité des fonctions elliptiques, Huzard-Courcier, t.1, Paris, 1825.
34. T. Leybourn, The Mathematical Questions Posed in the Ladies Diary, 1704–1816 (4 volumes), J. Mawman, London, 1817.
35. G. Lindner, Lexikon der gesamten Technik und ihrer Hilfswissenschaften, Band 3, second ed., Deutsche Verlagsanstalt, Stuttgart, 1904.
36. C. MacLaurin, A Treatise of Fluxions in Two Books, vol. 2, T. W. and T. Ruddimans, Edinburgh, 1742.
37. G. Mittag-Leffler, An introduction to the theory of elliptic functions, *Ann. of Math.*, 24 (1923) 271–351.
38. T. Muir, On the perimeter of an ellipse, *Messenger Math.*, 12 (1883) 149–151.
39. D. J. Newman, Rational approximation versus fast computer methods, Lectures on Approximation, and Value Distribution, Presses de l'Université de Montréal, 1982, pp. 149–174.
40. D. J. Newman, A simplified version of the fast algorithms of Brent and Salamin, *Math. Comp.*, 44 (1985) 207–210.
41. M. Nyvoll, Tilnaermelseformler for ellipsebuer, *Nordisk Mat. Tidskr.*, 25–26 (1978) 70–72.
42. G. Peano, Sur une formule d'approximation pour la rectification de l'ellipse, *C. R. Acad. Sci. Paris*, 108 (1889) 960–961.
43. T. Perl, The Ladies' Diary or Woman's Almanack, 1704–1841, *Historia Math.*, 6 (1979) 36–53.
44. S. Ramanujan, Modular equations and approximations to π, *Quart. J. Math.* (Oxford), 45 (1914) 350–372.
45. S. Ramanujan, Collected Papers, Chelsea, New York, 1962.
46. S. Ramanujan, Notebooks (2 volumes), Tata Institute of Fundamental Research, Bombay, 1957.
47. E. Salamin, Computation of π using arithmetic-geometric mean, *Math. Comp.*, 30 (1976) 565–570.
48. Science and the citizen, A bigger π, *Scientific American*, 248 (1983) 66.
49. E. S. Selmer, Bemerkninger til en ellipse-beregning av en ellipses omkrets, *Nordisk Mat. Tidskr.*, 23 (1975) 55–58.

50. Society of Mathematicians, The Diarian Repository or Mathematical Register, G. Robinson, London, 1774.
51. J. O. Stubban, Fergestads formel for tilnaermet beregning av en ellipses omkrets, *Nordisk Mat. Tidskr.*, 23 (1975) 51–54.
52. G. N. Watson, The marquis and the land-agent; a tale of the eighteenth century, *Math. Gaz.*, 17 (1933) 5–17.
53. E. T. Whittaker and G. N. Watson, A Course of Modern Analysis, 4th ed., Cambridge University Press, Cambridge, 1966.
54. J. Woyciechowsky, Sipos Pál egy kézirata és a kochleoid, *Mat. Fiz. Lapok*, 41 (1934) 45–54.

The Sequence of Pedal Triangles

JOHN G. KINGSTON AND JOHN L. SYNGE

JOHN G. KINGSTON has been at Nottingham University, in the Mathematics Department, since 1966, apart from temporary appointments in the U.S.A. and Canada.



JOHN L. SYNGE was born in Ireland in 1897 and has been a professor in Trinity College Dublin, University of Toronto, Ohio State University, Carnegie Institute of Technology, and the Dublin Institute for Advanced Studies, now emeritus.

Introduction. Although geometers have studied the properties of triangles for over two thousand years, there still remain problems of interest. For example, a given triangle T generates a sequence of triangles $\{T_n\}$ where T_{n+1} is the pedal triangle of T_n ($n \geqslant 0$, $T_0 = T$), this being the triangle whose vertices are the feet of the altitudes of T_n. This sequence was discussed by Hobson [4], [5] but, while his formulae for the transition from T_n to T_{n+1} are correct, those for T_n in terms of T are not. Lacking correct formulae, we experimented numerically, taking the angles of T to be integers in degrees. To our surprise the angles in the pedal sequence became periodic with periods of twelve steps. This curious fact led to an investigation of pedal sequences for general triangles, revealing that (a) the sequence may stop by degeneration of the triangle to a straight segment, (b) the angles may develop any periodicity, or (c) the sequence may proceed to infinity without periodicity.

In this article (Theorem I) we give necessary and sufficient conditions on the angles of T corresponding to these options. We show that any combination of angles which are rational multiples of π leads either to a periodic sequence or to degeneracy and obtain (Theorem II) an explicit expression for a period corresponding to these angles in terms of the Euler function ϕ.

A second surprising result which came to light, while we were experimenting with these cyclic sequences, was that eventually triangles became not only similar but also parallel to earlier triangles in the sequence. Indeed for an n-cycle in which T_n is similar to T, T_{n+1} is similar to T_1, etc, the triangle T_{6n} is both similar and also parallel to T (Theorem IV).

1. The pedal triangle. The pedal (or first pedal) triangle T' of a given triangle T of non-zero area is formed by joining the feet of the perpendiculars dropped from the vertices of T on the opposite sides, produced if necessary. Coxeter **[1]** prefers the word *orthic* to *pedal*, but we shall use the latter word, following a well-established tradition.

The pedal of T' is the second pedal of T, and so on. The sequence stops if we encounter a right-angled triangle; its pedal is a straight line segment. The condition for this is given in §3.

Hobson **[4]**, **[5]** discussed pedal sequences, stating correctly that the sides a', b', and c' and angles A', B' and C' of T' are given in terms of the elements of T by the following formulae: if A, B and C are all acute,

$$\begin{aligned} a' &= a\cos A, & b' &= b\cos B, & c' &= c\cos C; \\ A' &= \pi - 2A, & B' &= \pi - 2B, & C' &= \pi - 2C; \end{aligned} \tag{1.1}$$

if A is obtuse,

$$\begin{aligned} a' &= -a\cos A, & b' &= b\cos B, & c' &= c\cos C, \\ A' &= 2A - \pi, & B' &= 2B, & C' &= 2C, \end{aligned} \tag{1.2}$$

with similar formulae if B or C is obtuse. He gave formulae for the elements of the nth pedal, different according as n is odd or even. For n odd, his formula for the first angle of the nth pedal is

$$\tfrac{1}{3}(2^n + 1)\pi - 2^n A; \tag{1.3}$$

if $n = 1$ (first pedal) this gives $\pi - 2A$, correct for A acute by (1.1) but false for A obtuse by (1.2). Consequently Hobson's formulae for the angles of the nth pedal must be rejected. In fact the occurrence of obtuse angles makes it impossible to write down any reasonably simple explicit formulae for the angles of the nth pedal. But Hobson's formulae for the sides of the nth pedal are correct to within a sign, and we may write for the first side of the nth pedal

$$a_n = \pm a\cos A\cos 2A\cos 4A \cdots \cos 2^{n-1}A, \tag{1.4}$$

the sign being chosen to give a positive value.

If we write $A = \alpha\pi$, $B = \beta\pi$, $C = \gamma\pi$, then $\alpha + \beta + \gamma = 1$ and (α, β, γ) can be interpreted as the barycentric coordinates of a point interior to a triangle of reference PQR (see Figure 1). The formulae given in (1.1) and (1.2) for the angles of the pedal triangle then have a particularly pleasing representation. If A is obtuse, α is greater than $\frac{1}{2}$ and (α, β, γ), which represents the angles of T, maps to the point $(2\alpha - 1, 2\beta, 2\gamma)$, which represents the angles of the pedal triangle T'. In the triangle PQR this is a dilatation from P $(1, 0, 0)$ with scale factor 2. Similarly for B or C obtuse we have corresponding dilatations from Q or R. If the triangle is acute-angled then $\alpha, \beta, \gamma < 1/2$ and (1.1) shows that (α, β, γ) maps to $(1 - 2\alpha, 1 - 2\beta, 1 - 2\gamma)$. This is a dilatation from S, the point $(1/3, 1/3, 1/3)$ with scale factor -2. We thus have a discontinuous (when one of α, β, γ is $1/2$) 4-to-1 mapping of the triangle PQR to itself.

As an example we consider a triangle T with angles represented by barycentric coordinates $(\frac{40}{127}, \frac{43}{127}, \frac{44}{127})$. Figure 1 shows the locations of the points which represent T and its subsequent pedal triangles $T_1, T_2, \ldots, T_7$. Note that $T_7 = T$. This is the sequence of triangles (5.3) which arises in §5.

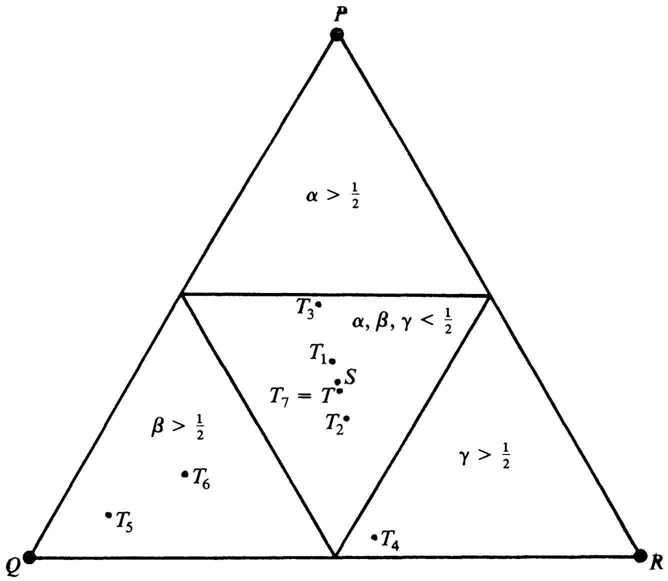


FIG. 1

2. The angles of the *n*th pedal triangle. The notation A^i, a^i, $i = 1, 2, 3$ is now introduced to represent the angles A, B, C and sides a, b, c of the triangle T. The angles and sides of the nth pedal triangle T_n will thus be represented as A_n^i and a_n^i. Equations (1.1) and (1.2) show that the angles progress according to the rule

$$A_1^i \equiv \pm 2A^i (\operatorname{mod} \pi), \qquad i = 1, 2, 3, \tag{2.1}$$

where the sign depends on the nature of the original triangle T, but is the same for each i. For the angles of T_n,

$$A_n^i \equiv 2^n E_n A^i (\operatorname{mod} \pi), \tag{2.2}$$

where $E_n = \pm 1$ and is independent of i.

3. Pedal degeneracy. The pedal of a right-angled triangle is a straight segment and this has no pedal; the pedal sequence stops. We call T pedally degenerate (PD) iff a right angle occurs in T or in the pedal sequence which starts from T. It is evident from (2.2) that T is PD iff one of its angles is of the form

$$\pi(2m - 1)/2^n, \tag{3.1}$$

where m and n are positive integers. Thus the angles responsible for pedal degeneracy are

$$\tfrac{1}{2}\pi,\ \tfrac{1}{4}\pi,\ \tfrac{1}{8}\pi \dots\ \tfrac{3}{4}\pi,\ \tfrac{3}{8}\pi, \dots,\ \tfrac{5}{8}\pi,\ \tfrac{5}{16}\pi, \dots .$$

A pedal sequence is infinite if it starts from a triangle which is not PD.

4. Pedal cycles. Let T_s and T_{s+n} $(n > 0)$ be triangles in the pedal sequence which starts with T assumed not PD. If $A_s^i = A_{s+n}^i$, $i = 1, 2, 3$ (that is corresponding angles of T_s and T_{s+n} are equal) we say that we have an *n-cycle* starting with T_s. It is clear that T_s is similar to $T_{s+n}, T_{s+2n}, \dots$.

We now describe the nature of the angles of T which give rise to an n-cycle by means of the following theorem.

THEOREM I. *If $s \geqslant 0$, $n > 0$ and T is not PD, then T_s starts an n-cycle iff the angles of T are of the form*

$$A^i = \pi p^i/(2^s(2^n + E)) \qquad i = 1,2,3, \tag{4.1}$$

where $E = \pm 1$ and p^i are positive integers satisfying

$$p^1 + p^2 + p^3 = 2^s(2^n + E). \tag{4.2}$$

Proof of sufficiency. We have to show that, if the angles of T are of the form (4.1) then $A^i_{s+n} = A^i_s$. Using (2.2) with (4.1) gives

$$A^i_s \equiv \pm\pi p^i/(2^n + E) \pmod{\pi}$$

and

$$A^i_{s+n} \equiv \pm\pi p^i 2^n/(2^n + E) \pmod{\pi}.$$

The signs in these two equations may differ but for each equation the same sign applies for all i. In any of the 4 cases either $A^i_s + A^i_{s+n}$ or $A^i_s - A^i_{s+n}$ is an integral multiple of π. Since A^i_s and A^i_{s+n} are both in the interval $(0, \pi)$ they are therefore either equal or supplementary, and in either case have the same sine. By the sine rule the ratios of the lengths of a triangle are the same as the ratios of the sines of the opposite angles. The triangles T_s and T_{s+n}, therefore, have their sides in the same ratio, and are similar. We must also have $A^i_{s+n} = A^i_s$.

Proof of necessity. We have to show that if $A^i_{s+n} = A^i_s$, then the angles of T are as in (4.1). Equation (2.2) gives

$$A^i_s \equiv 2^s E_s A^i \pmod{\pi}, \qquad A^i_{s+n} \equiv 2^{s+n} E_{s+n} A^i \pmod{\pi},$$

where $E_s = \pm 1$, $E_{s+n} = \pm 1$. Combining these two equations gives

$$2^s(2^n + E)A^i \equiv 0 \pmod{\pi},$$

where $E = \pm 1$ and is independent of i, so that A^i are indeed of the form (4.1). Necessity is now proved.

Note that in Theorem I the combination $n = 1$ and $E = -1$ may be excluded since T is then PD. This means that 1-cycles only occur when $E = +1$. For $n > 1$ n-cycles exist whether $E = +1$ or $E = -1$.

It might appear from the preceding theorem that, if the angles of T are as in (4.1), then the n-cycle does not start until we reach T_s. But that is not necessarily the case. If the values of p^i in the partition happen to have the common factor 2^j $(1 \leqslant j \leqslant s)$ then $p^i = 2^j q^i$ and (4.1) yields $A^i = \pi q^i/(2^{s-j}(2^n + E))$. Applying Theorem I we now see that $A^i_{s-j} = A^i_{s-j+n}$, so that the n-cycle begins with T_{s-j} (if not earlier) for such choices of the p^i.

5. Examples. The only triangle which maintains its form under the pedal process is the equilateral triangle: here we have a monocycle ($n = 1$), and it seems that there is no more to be said. But Theorem I shows that we may start with a scalene triangle which, after a possible delay represented by s, settles down into a monocycle. Choose, for example, $n = 1$, $s = 4$, $E = 1$, so that $2^s(2^n + E) = 48$. We are now to take a partition of 48, but taking care to avoid pedal degeneracy. Thus $48 = 3 + 5$

$+ 40$ and, since $3/48 = 1/16$, (3.1) shows that we have a PD triangle. But $48 = 4 + 7 + 37$ and the corresponding triangle is not PD. To see how the sequence develops, we go back to (1.1) and (1.2), starting with the angles $\frac{4}{48}\pi$, $\frac{7}{48}\pi$, $\frac{37}{48}\pi$. The factor $\frac{1}{48}\pi$ being understood, so that a right angle is represented by 24, and with obtuse angles marked by an asterisk, the calculation proceeds as follows:

$$\begin{array}{ccc} 4 & 7 & 37^* \\ 8 & 14 & 26^* \\ 16 & 28^* & 4 \\ 32^* & 8 & 8 \\ 16 & 16 & 16 \\ 16 & 16 & 16 \end{array} \tag{5.1}$$

The monocycle has been established in the fourth step.

In §6 we shall discuss the dodekacycle ($n = 12$) from which this work originated. As further examples here we shall consider two heptacycles ($n = 7$) without delay ($s = 0$), one corresponding to each value of E.

For $E = 1$, we have $2^s(2^n + E) = 129$, and we see from (3.1) that no partition can represent a triangle which is PD. Let us see what happens to a triangle which is nearly equilateral, displaying the sequence as before (this time with a factor of $\frac{1}{129}\pi$ being understood and a right angle being represented by $64\frac{1}{2}$).

$$\begin{array}{ccc} 42 & 43 & 44 \\ 45 & 43 & 41 \\ 39 & 43 & 47 \\ 51 & 43 & 35 \\ 27 & 43 & 59 \\ 75^* & 43 & 11 \\ 21 & 86^* & 22 \\ 42 & 43 & 44 \end{array} \tag{5.2}$$

The cycle is completed in seven steps.

For $E = -1$, we have $2^s(2^n + E) = 127$, which is prime, and no partition can give a triangle which is PD. With a right angle at $63\frac{1}{2}$, and a triangle nearly equilateral, we have

$$\begin{array}{ccc} 40 & 43 & 44 \\ 47 & 41 & 39 \\ 33 & 45 & 49 \\ 61 & 37 & 29 \\ 5 & 53 & 69^* \\ 10 & 106^* & 11 \\ 20 & 85^* & 22 \\ 40 & 43 & 44 \end{array} \tag{5.3}$$

The cycle is completed in seven steps.

6. Dodekacycles. The preceding results are the outcome of observing the existence of 12-cycles in numerical experiments with triangles whose angles were each an integral number of degrees. This can be proved by noting that $2^2(2^{12} - 1)$ is a multiple of 180, so that each such angle can be represented in the form (4.1) with

$n = 12$, $s = 2$ and $E = -1$. By Theorem I, a 12-cycle begins after a delay of at most two steps. Pedal degeneracy is avoided if the triangle does not have an angle of 45°, 90° or 135°.

To illustrate the delays, here is an example, the angles being given in degrees:

61	63	56
58	54	68
64	72	44
52	36	92*
104*	72	4
28	144*	8
56	108*	16
112*	36	32
44	72	64
92*	36	52
4	72	104*
8	144*	28
16	108*	56
32	36	112*
64	72	44

We see that, depending on whether we start with T having angles (61, 63, 56), (58, 54, 68) or (64, 72, 44) the 12-cycle starts with T_2, T_1 or T, respectively.

7. Cycle lengths. We have to thank the Babylonians for dividing the semicircle into 180 degrees, since otherwise pedal cycles might have remained undiscovered. But a closer examination shows that cycles would have been found had we taken angles integral in seconds of arc or integral in grades (100 grades = 90°) or, more generally (in radians), as any rational multiples of π,

$$A^i = \pi q^i / M, \tag{7.1}$$

where q^i and M are positive integers.

The question now arises of how long the cycles will be for a given initial set of angles. A periodic length will be determined (Theorem II) in this section but the result may not give the shortest period. For example, we may prove that a 12-cycle exists when it may be the case that 6-cycles exist.

By Theorem I we know that, if T has angles given by (7.1), and is not PD, it generates an n-cycle if there exist integers $p^i > 0$, $s \geqslant 0$, $n > 0$ to satisfy

$$q^i 2^s (2^n + E) = p^i M, \tag{7.2}$$

with $E = 1$ or -1.

A solution (not unique) can always be found by making use of the Euler function ϕ whose relevant properties (Dickson [2]) are as follows: $\phi(m)$ is the number of positive integers, not exceeding the positive integer m, which are relatively prime to m: if $p_1, p_2, \ldots, p_k$ are the distinct prime factors of m, then

$$\phi(m) = m(1 - p_1^{-1})(1 - p_2^{-1}) \cdots (1 - p_k^{-1}); \tag{7.3}$$

if a is prime to m,

$$a^{\phi(m)} \equiv 1 \pmod{m}. \tag{7.4}$$

THEOREM II. *If a triangle T is not PD and has angles*

$$A^i = \pi q^i/M,$$

where q^i, M are positive integers, then T generates an n-cycle with $n = \frac{1}{2}\phi(M_1)$ where M_1 is the odd integer given by $2^t M_1 = M$, where t is zero or a positive integer.

Proof. Since M_1 is odd and T is not PD it follows that $M_1 \geqslant 3$. Hence the form of $\phi(M_1)$ given in (7.3) is even and it may be shown that

$$2^{\phi(M_1)/2} \equiv E_1 \pmod{M_1},$$

where $E_1 = 1$ or -1 (see Appendix). A solution of (7.2) may be obtained by taking

$$s = t, \qquad E = -E_1, \qquad n = \phi(M_1)/2$$

$$p^i = \gamma q^i,$$

where γ is the positive integer determined by

$$2^{\phi(M_1)/2} - E_1 = \gamma M_1,$$

and so the theorem is proved.

Let us apply the theorem to the case of a triangle with angles integral in degrees. If $M = 180 = 2^2 \times 45$, so that $s = 2$, $M_1 = 45$, $\phi(M_1) = 24$, Theorem II shows that we must get 12-cycles as already proved in §6. For $M = 180$ Theorem II gives the "best" result in that triangles exist whose shortest pedal cycle is a 12-cycle.

We invite the reader to experiment with triangles whose angles are integral in grades, for which Theorem II predicts 10-cycles.

Theorem II does not always give the best possible result and the value $\phi(M_1)/2$ for cycle length can be wildly out, compared with the shortest cycle, as the following example (Ribenboim [6]) shows: when $M = M_1 = 1093^2$ (1093 is prime), $\phi(M_1)/2 = 1093 \times 546 = 596{,}778$. But $2^{182} \equiv -1 \pmod{M_1}$, so that a solution of (7.2) may be found with $n = 182$, and we may conclude that there exists a cycle of length 182, a considerable improvement on $\phi(M_1)/2$.

8. A projective map. A well-known property of elementary geometry (see, for example, Theorem 20 in Durell [3]) is that if the pedal triangle T' of a triangle T is magnified by a factor 2 with respect to the centre H, the orthocentre of T, then the vertices of the magnified T' will lie on the circumcircle of T. Hence, if this magnification is performed at each stage of the pedal sequence, we may represent all members of a pedal sequence by triads of points on a single circle, the circumcircle of the triangle from which the sequence starts. This is illustrated in Figure 2. For future use we introduce some further notation. After n of these pedal maps we suppose that the vertices of A, B, C of triangle T are mapped to $\bar{A}_n$, $\bar{B}_n$, $\bar{C}_n$ of triangle $\bar{T}_n$. The properties of the pedal map show that $\bar{T}_n$ will be similar and parallel to the actual nth pedal triangle T_n but magnified by factor 2^n.

9. Pedal ancestry. Given a triangle T which is not PD, it generates a unique infinite pedal sequence, which may be cyclic or, in general, not so. A triangle T_{-1} of which T is the pedal may be called an *antipedal* of T and so we are led to consider

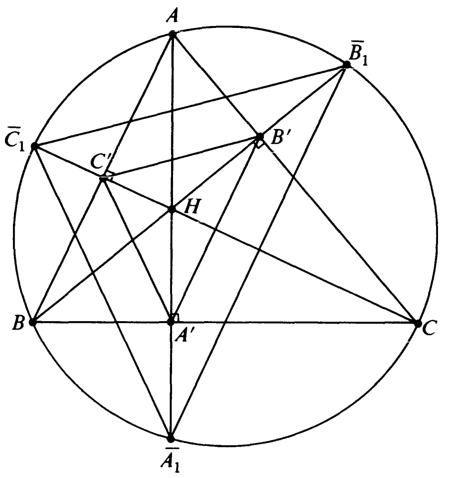


FIG. 2(a).

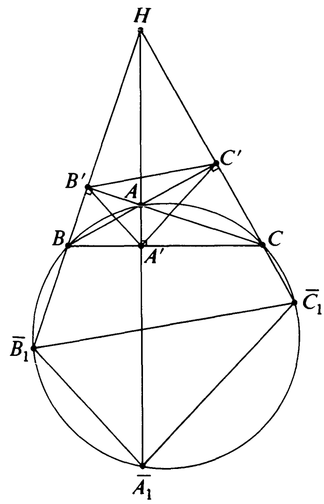


FIG. 2(b).

all triangles T_{-n} (with n positive) which include T in the pedal sequence generated by them. To use a biological term, these constitute the *ancestry* of T.

The relations between the angles A_n^i, A_{n+1}^i of the triangles T_n and T_{n+1} may be obtained from (1.1) and (1.2). If T_n is acute-angled then (1.1) gives $A_{n+1}^i = \pi - 2A_n^i$, so that

$$A_n^i = \tfrac{1}{2}\left(\pi - A_{n+1}^i\right), \qquad i = 1, 2, 3. \tag{9.1}$$

If A_n^1 is an obtuse angle then (1.2) gives $A_{n+1}^1 = 2A_n^1 - \pi$, $A_{n+1}^2 = 2A_n^2$, $A_{n+1}^3 = 2A_n^3$, so that

$$A_n^1 = \tfrac{1}{2}\left(A_{n+1}^1 + \pi\right), \qquad A_n^2 = \tfrac{1}{2}A_{n+1}^2, \qquad A_n^3 = \tfrac{1}{2}A_{n+1}^3. \tag{9.2}$$

There are two other possibilities, namely A_n^2 obtuse and A_n^3 obtuse. For these cases formulae analogous to (9.2) hold.

There is thus a fourfold ambiguity in the antipedal: any triangle has four parents, each step backward in the pedal sequence being determined only by our decision to make T_n acute-angled, or to have a specified angle obtuse.

In terms of the representation in barycentric coordinates illustrated in Figure 1, for each triangle represented by a point within PQR there is one antipedal in each of the 4 triangular regions shown within PQR.

Every person now living has had a unique ancestor a thousand generations ago if one goes back along a male line or a female line. Similarly, every triangle has a unique line of ancestors if it proceeds through acute-angled triangles or through triangles in which some specified angle is obtuse.

To follow the acute ancestry we note that (9.1) can be arranged

$$A_n^i - \frac{\pi}{3} = -\frac{1}{2}\left(A_{n+1}^i - \frac{\pi}{3}\right). \tag{9.3}$$

Thus for any triangle, acute or obtuse, we see that as we trace back this line of ancestry the angles of successive triangles converge to $\pi/3$. The triangles however rapidly become very large since the circumradius is doubled in each step backward.

As for the obtuse ancestry, with A_n^1 obtuse say, we use (9.2). Thus at each step we halve the angles A_n^2 and A_n^3 and see that the angles in the sequence of ancestors converge to π, 0, and 0.

The fourfold ambiguity in the antipedal renders the backward extension of a pedal cycle complicated, since, to reproduce it, we would need to choose the proper parent at each step. Our knowledge of the distribution of obtuse angles in a cycle is so far purely experimental. But note that no cycle, except the 1-cycle, can consist of purely acute angles, for (9.3) shows that, as a sequence of acute triangles progresses, each acute angle differs from $\pi/3$ by twice as much as its predecessor. Obviously one angle will sooner or later become obtuse.

10. Spin and rotation in a pedal sequence. All of the properties of pedal sequences which have been discussed so far have referred only to the shape and size of the triangles. We now investigate some of the positional properties of such sequences of triangles.

By the *spin* of a triangle $A_nB_nC_n$ in a pedal sequence we mean the counterclockwise or clockwise arrangements of the vertices, taken in alphabetical order. Note that, if a triangle is acute-angled (Figure (2a)), the spin of its pedal triangle is the same as that of the original triangle whereas if a triangle has an obtuse angle (Figure 2(b)), the spin of the pedal triangle is reversed from that of the original triangle.

In order to obtain the main results of this section, Theorems III and IV, we now prove a preliminary result concerning the rotation of the sides of a triangle in one stage in the pedal sequence.

LEMMA. *Let triangle ABC have counterclockwise spin and let A', B', C' be the corresponding vertices of its pedal triangle. Let $R(BC)$, $R(CA)$, and $R(AB)$ be the angles through which the directed sides BC, CA, and AB must be turned counterclockwise to make them directed respectively the same as the directed sides $B'C'$, $C'A'$, $A'B'$. Then*

$$R(BC) \equiv \begin{cases} B - C + \pi & (\text{if } A \text{ is acute}) \\ B - C & (\text{if } A \text{ is obtuse}) \end{cases},$$

where the congruence is modulo 2π. Corresponding expressions hold for $R(CA)$ and $R(AB)$.

To verify the lemma for the case where all angles of triangle ABC are acute, refer to FIGURE 2(a). If we turn BC to BB', then to $B'B$, and finally to $B'C'$, and accumulate the counterclockwise angle of turning between each successive pair of directed segments, we obtain, modulo 2π,

$$R(BC) \equiv \left(\frac{\pi}{2} - C\right) + \pi - \left(\frac{\pi}{2} - B\right) = B - C + \pi.$$

(To see that angle $BB'C'$ is equal to $\pi/2 - B$, note that both B' and C' lie on the circle having diameter BC. Angles $BB'C'$ and BCC' are inscribed angles of this circle which include the same arc BC', and are thus equal. But angle BCC' is the complement of angle B). The case where A is obtuse is shown in Figure 2(b), and the case where either angle B or angle C is obtuse is not illustrated, but in either of these cases it is straightforward to verify the lemma, using the same approach as above.

THEOREM III. *If triangle ABC has pedal triangle $A'B'C'$, then, using the notation of the preceding lemma,*

$$R(BC) + R(CA) + R(AB) \equiv \begin{cases} \pi & (\text{if all angles of } ABC \text{ are acute}) \\ 0 & (\text{if } ABC \text{ has an obtuse angle}) \end{cases},$$

where the congruence is modulo 2π.

This is proved, if ABC has counterclockwise spin, by combining the expressions which the lemma gives for $R(BC)$, $R(CA)$ and $R(AB)$. If ABC has clockwise spin, the values of $R(BC)$, $R(CA)$ and $R(AB)$, in the lemma, have opposite sign and the conclusion of the theorem, being modulo 2π, still holds.

THEOREM IV. *If triangle T starts an n-cycle then the triangle T_{6n}, obtained after six complete cycles, is oriented similarly to T and, if T_{6n} is magnified by the factor 2^{6n} with respect to any point, the resulting triangle may be obtained from T by giving T a translation.*

This theorem is proved by considering two cases.

Case 1. T_n has spin opposite to that of T. Clearly the triangles encountered in the continuation of the pedal sequence from T_n to T_{2n} will be similar to those in the cycle from T to T_n, but with reversed spins. All of the angles of turning for the sides, involved in the steps between T_n and T_{2n} will be the same as those in the steps between T and T_n, except with signs reversed, so that the net angle of turning for each side during this double-length cycle is zero. Then T_{2n} is oriented the same as T, and of course T_{6n} is also.

Case 2. T_n has the same spin as T. If T has vertices A, B, C let $R_n(BC)$ denote the total counterclockwise rotation of the directed side BC (and its successors) in the steps of the cycle from T to T_n, and similarly for $R_n(CA)$ and $R_n(AB)$. Since T_n is similar to T, with the same spin, we have

$$R_n(BC) \equiv R_n(CA) \equiv R_n(AB) \pmod{2\pi}.$$

Also, $R_n(BC) + R_n(CA) + R_n(AB)$ can be found by adding the expressions given by Theorem III for each step of the n-cycle, each of which is 0 or π, and so this total is an integral multiple of π. Therefore $R_n(BC)$, $R_n(CA)$, $R_n(AB)$ are integral multiples of $\pi/3$. Since T_n has the same spin as T, the turning of the sides in successive n-cycles is added on in the same sense (instead of cancelling as in case 1). Thus in six complete cycles each side will have been rotated through an integral number of complete rotations.

In either case, T_{6n} is oriented the same as T, with correspondingly lettered vertices in corresponding positions. The discussion in §8 shows that if T_{6n} is magnified by the factor 2^{6n} it can be translated to coincide with T.

To illustrate the preceding results let us take $s = 0$, $n = 4$, $E = -1$ in (4.2), so that $2^s(2^n + E) = 15$, and start with the partition $(1, 3, 11)$. Then, with a factor $\pi/15$ understood, the 4-cycle is as follows:

1	3	11*
2	6	7
11*	3	1
7	6	2
1	3	11*

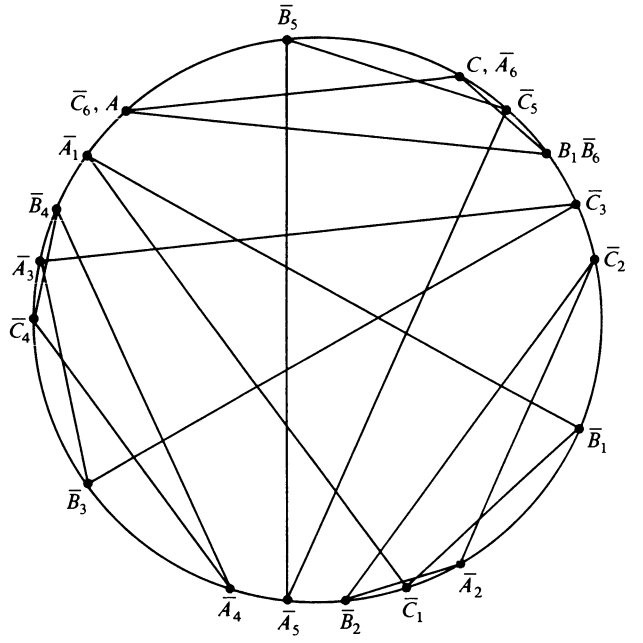


FIG. 3.

FIGURE 3 illustrates this pedal sequence as far as the sixth step, repeatedly using the pedal map described in §8 to magnify the triangles so that they are all inscribed in the same circle. (This avoids microscopically small triangles.) Note that $\bar{T}_6$ (which is T_6 magnified) in fact coincides with T, but with $\bar{A}_6 = C$ and $\bar{C}_6 = A$. Evidently another 6 steps will produce $\bar{T}_{12}$ in exactly the same position as T with $\bar{A}_{12} = A$, $\bar{B}_{12} = B$, $\bar{C}_{12} = C$. Theorem IV predicts that six 4-cycles produce a triangle T_{24} which when magnified by 2^{24} will be a pure translation of T. The fact that in this example we only need three 4-cycles to make analogous claims for T_{12} is because the basic 4-cycle has an even number (two) of acute-angled triangles making $R_n(BC)$, $R_n(CA)$, $R_n(AB)$ *even* multiples of $\pi/3$ in case 2 (T_4 has the same spin as T) of the proof of Theorem IV.

11. Conclusion. We have discovered two facts: the negative fact that Hobson was in error and the positive fact that pedal cycles exist. In general we have been unable to locate the *pedal point*, by which we mean the point which is the limit of an infinite pedal sequence, cyclic or not. However, one consequence of Theorem IV is that if a triangle starts a pedal n-cycle then its pedal point is collinear with A and A_{6n}, with B and B_{6n} and with C and C_{6n}.

Professor Synge thanks Professor J. T. Lewis, a colleague, for useful discussions. We are also grateful to the referees for their helpful comments and acknowledge the use of their suggestions.

Appendix

THEOREM. *If $n \geqslant 3$ is an odd integer, then*

$$2^{\phi(n)/2} \equiv \pm 1 \pmod{n} \tag{A.1}$$

Proof.

Case 1. n is a power of a single prime factor.

Suppose $n = p^s$ where $p \geqslant 3$ is prime and where $s \geqslant 1$. Then, since 2 is prime to n, (7.4) gives

$$2^{\phi(n)} - 1 \equiv 0 \pmod{p^s}. \tag{A.2}$$

(7.3) shows that $\phi(n)$ is even and (A.2) may, therefore, be factorized to give

$$(2^{\phi(n)/2} + 1)(2^{\phi(n)/2} - 1) \equiv 0 \pmod{p^s}. \tag{A.3}$$

The two factors on the left-hand side of (A.3) differ by 2 and, since $p \geqslant 3$, they cannot both have p as a factor. Hence one of the factors must be divisible by p^s and (A.1) immediately follows.

Case 2. n has two or more different primes as factors.

Suppose $n = p^sQ$ where p is prime, $s \geqslant 1$, and where $Q \geqslant 3$ is not divisible by 2 or p. If $q_1, \ldots, q_j$ are the distinct prime factors of Q, we have from (7.3) that

$$\phi(n) = p^{s-1}(p-1)Q\left(1 - q_1^{-1}\right) \cdots \left(1 - q_j^{-1}\right). \tag{A.4}$$

Also from (7.3),

$$\phi(p^s) = p^{s-1}(p-1). \tag{A.5}$$

Comparing (A.4) and (A.5) it is clear that, since $j \geqslant 1$, $\phi(n)$ is an even multiple of $\phi(p^s)$. But case 1 gives

$$2^{\phi(p^s)/2} \equiv \pm 1 \pmod{p^s},$$

so it follows that

$$2^{\phi(n)/2} \equiv 1 \pmod{p^s}.$$

Since this last congruence holds for each prime factor of n in place of p (with s depending on p), the congruence must hold modulo n, giving (A.1) and completing the proof.

REFERENCES

1 H. S. M. Coxeter, Introduction to Geometry, John Wiley and Sons, New York, 1969, pp. 5, 18.
2 L. E. Dickson, Modern Elementary Theory of Numbers, Chicago University Press, 1939, pp. 9-12.
3 C. V. Durell, Modern Geometry, McMillan and Co., London, 1952, p. 31.
4 E. W. Hobson, A Treatise on Plane Trigonometry, Cambridge University Press, 1897, pp. 194-200.
5 ______, A Treatise on Plane Trigonometry, Cambridge University Press, 1925, pp. 197-203.
6 P. Ribenboim, Algebraic Numbers, Wiley- Interscience, New York, 1972, p. 64.

Magic Squares Over Fields*

CHARLES SMALL, QUEEN'S UNIVERSITY

1. Introduction

A magic square of size n over a field F is an $n \times n$ matrix with entries in F for which every row, every column, the principal diagonal, and the principal backdiagonal all have the same sum. The set of all such matrices is an F-vector space. We compute its dimension. The answer is $n^2 - 2n$ (independent of F) for $n \geq 5$; for $n < 5$ the results depend on the characteristic of F.

Magic squares *with integer entries* have a rich history and an extensive literature: see, for example, section A20 of **[1]** for an indication of recent work. Magic squares *over fields* have been comparatively neglected, however, and throughout the literature the emphasis is on generating new magic squares rather than examining the structure of all magic squares of fixed size. In **[2]** a theorem equivalent to ours is stated for magic squares of size $n \geq 3$ over the real numbers.

While the present results are, therefore, probably not new, except possibly in the case of positive characteristic, the proofs may be worth consideration as an appealing application of elementary techniques in linear algebra. Two features are worthy of special note (pedagogically as well as mathematically): the phenomenon, widespread in mathematics and here particularly transparent in the proof, of peculiarities in low dimensions which vanish as soon as there is "enough room"; and the special role, in the low-dimensional cases, of positive characteristic.

To begin, we make the definition precise: a *magic square of size n and weight d* over a field F is a matrix $A = (a_{ij})$ $(1 \leq i, j \leq n)$ with entries in F satisfying:

$$(*) \quad \begin{cases} \displaystyle\sum_{i=1}^{n} a_{ij} = d & \text{for all } j \\ \displaystyle\sum_{j=1}^{n} a_{ij} = d & \text{for all } i \\ \displaystyle\sum_{i=1}^{n} a_{ii} = d & \\ \displaystyle\sum_{i=1}^{n} a_{i,n-i+1} = d. & \end{cases}$$

Here $n > 1$ is an integer and $d \in F$.

Let B be the transpose of the row vector (with n^2 entries) obtained by listing the rows of A in order:

$$B = (a_{11}, a_{12}, \ldots, a_{1n}, a_{21}, \ldots, a_{2n}, \ldots, a_{n1}, \ldots, a_{nn})^t.$$

CHARLES SMALL received his Ph.D. in 1969 from Columbia University, and has been at Queen's University since 1970. He is coauthor of books on the Brauer group and on homological methods in commutative algebra, and has published 20 papers in algebra and number theory, two with coauthors. Current research interests include permutation polynomials and other aspects of arithmetic in finite fields.

*Acknowledgement. Thanks to R. D. Pollack, who asked the question.

Then clearly the $2n + 2$ conditions (*) can be written as a single matrix equation $C_n B = d(1,1,\ldots,1)^t$, where C_n is the $(2n+2) \times n^2$ matrix constructed as follows:

Let $\underline{e}_1 = (1,0,\ldots,0)$, $\underline{e}_2 = (0,1,0,\ldots,0),\ldots,$ $\underline{e}_n = (0,\ldots,0,1)$ be the standard basis for F^n and let

$$\underline{j} = (1,1,\ldots,1) = \sum_{i=1}^{n} \underline{e}_i.$$

The first row of C_n is $(\underline{e}_1, \underline{e}_1, \ldots, \underline{e}_1)$, the second is $(\underline{e}_2, \underline{e}_2, \ldots, \underline{e}_2)$, and so on to get the first n rows. The $(n+1)$st row of C_n is $(\underline{j}, \underline{0}, \ldots, \underline{0})$, the $(n+2)$nd row is $(\underline{0}, \underline{j}, \underline{0}, \ldots, \underline{0})$, and so on to get the next n rows; here $\underline{0}$ is the n-tuple of 0's. Finally, the $(2n+1)$st row of C_n is $(\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_n)$ and the $(2n+2)$nd is $(\underline{e}_n, \underline{e}_{n-1}, \ldots, \underline{e}_1)$.

The set $M_n(F)$ of magic squares of size n (and arbitrary weight) over F is clearly an F-vector space with respect to the usual matrix operations; and the mapping $w: M_n(F) \to F$ which assigns to any magic square its weight is clearly linear. The kernel $Z_n(F)$ of w is the set of magic squares of size n and weight 0. We let $m_n(F) = \dim_F M_n(F)$ and $z_n(F) = \dim_F Z_n(F)$. There are two possibilities: either w is onto and $m_n(F) = z_n(F) + 1$, or w is the 0 map and $m_n(F) = z_n(F)$. We will see (§3, Corollary) that the latter situation arises only for $n = 2$ or 3 and $\operatorname{char.} F = n$. Notice that w is certainly onto in characteristic 0, and in characteristic p when $p \nmid n$, since in these cases the matrix with all entries 1 is a magic square of nonzero weight.

It is clear from this discussion that to compute $m_n(F)$ we must compute the rank of C_n and determine exactly when w is 0. In §3 we do these simultaneously by computing the rank of C_n augmented by a column of 1's. Since the argument there requires $n \geq 5$, we pause first to handle the anomalous cases $n \leq 4$.

2. The cases $n \leq 4$

THEOREM. (a) *For $n = 2$ we have $z_2(F) = 0$ and $m_2(F) = 1$, unless char. $F = 2$ in which case $z_2(F) = m_2(F) = 1$.*

(b) *For $n = 3$ we have $z_3(F) = 2$ and $m_3(F) = 3$, unless char. $F = 3$ in which case $z_3(F) = m_3(F) = 3$.*

(c) *For $n = 4$ we have $z_4(F) = 7$ and $m_4(F) = 8$, unless char. $F = 2$ in which case $z_4(F) = 8$ and $m_4(F) = 9$.*

Proof. (a) It is immediate that for any F the only magic squares of size 2 are the scalar multiples of

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

The results follow.

(b) For $n = 3$ we augment C_3 by a column of 1's and row-reduce to get

$$\left(\begin{array}{ccccccccc|c}
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & :1 \\
0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & :1 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & :1 \\
0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & :1 \\
0 & 0 & 0 & 0 & 1 & -1 & 1 & 0 & -1 & :0 \\
0 & 0 & 0 & 0 & 0 & 3 & -3 & 0 & 3 & :1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & :1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & :0
\end{array}\right).$$

Thus C_3 has rank 7 unless char. $F = 3$, in which case the rank is 6. Hence $z_3(F) = 9 - 7 = 2$ unless char. $F = 3$, in which case $z_3(F) = 9 - 6 = 3$. The augmenting column gives no trouble (and w is onto) unless char. $F = 3$, in which case there is trouble in the sixth row (and w is 0). The results follow.

(c) For $n = 4$ we augment C_4 by a column of 1's and row-reduce to get

$$\begin{pmatrix}
1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&:1\\
0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&:1\\
0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&:1\\
0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&:1\\
0&0&0&0&1&1&1&1&0&0&0&0&0&0&0&0&:1\\
0&0&0&0&0&2&1&1&-1&0&1&0&-1&0&0&1&:1\\
0&0&0&0&0&0&1&-1&0&1&0&-1&1&0&0&-1&:0\\
0&0&0&0&0&0&0&0&1&1&1&1&0&0&0&0&:1\\
0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1&:1\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&:0
\end{pmatrix}.$$

If char. $F \neq 2$ the rank is 9 and there is no trouble from the augmenting column (so w is onto), so that $z_4(F) = 16 - 9 = 7$ and $m_4(F) = 8$ in this case. In characteristic 2 the obvious additional row reductions show that the rank is 8 (with, again, no trouble from the augmenting column): the eighth row is now the sum of the sixth and seventh. The results follow.

Remark. It is an easy exercise, which we leave to the reader, to carry the row reductions of C_3 and C_4 a little further and then read off bases for $Z_3(F)$ and $Z_4(F)$, and (if the augmenting column of 1's is retained) for $M_3(F)$ and $M_4(F)$. The same remark applies for $n = 5, 6, \ldots$.

Examples:

(a)
$$\begin{pmatrix}1&0&1&1\\1&1&1&0\\1&1&0&1\\0&1&1&1\end{pmatrix}$$

is a magic square of size 4 and weight 1 in characteristic 2; so is

$$\begin{pmatrix}1+x&0&1+x&1\\1&1+x&1&x\\1&1+x&0&1+x\\x&1&1+x&1\end{pmatrix}\quad\text{for any } x.$$

(b)
$$\begin{pmatrix}1&1&1\\1&1&1\\1&1&1\end{pmatrix},\begin{pmatrix}1&-1&0\\0&1&-1\\-1&0&1\end{pmatrix},\begin{pmatrix}1&0&-1\\0&-1&1\\-1&1&0\end{pmatrix}$$

form a basis for the magic squares of size 3 in characteristic 3.

(c) In characteristic zero,

$$X = \begin{pmatrix}1&0&-1\\-2&0&2\\1&0&-1\end{pmatrix}\quad\text{and}\quad Y = \begin{pmatrix}0&1&-1\\-1&0&1\\1&-1&0\end{pmatrix}$$

form a basis for the magic squares of size 3 and weight 0, while X and Y together

with

$$Z = \begin{pmatrix} 2 & 2 & -1 \\ -2 & 1 & 4 \\ 3 & 0 & 0 \end{pmatrix}$$

form a basis for the space of all magic squares of size 3. Thus a typical 3×3 magic square in characteristic 0 has the form

$$\begin{pmatrix} 2r+s & 2r+t & -r-s-t \\ -2r-2s-t & r & 4r+2s+t \\ 3r+s+t & -t & -s \end{pmatrix};$$

its weight is $3r$.

3. The general case, $n \geq 5$

As in §2, we argue by augmenting C_n with a column of 1's and row reducing. We will see that assuming $n \geq 5$ gives enough room so that the case distinctions depending on char. F no longer arise.

THEOREM. *For $n \geq 5$, $m_n(F) = n(n-2) = z_n(F) + 1$ for all fields F.*

COROLLARY. *The weight mapping w is onto unless char. $F = n = 2$ or char. $F = n = 3$. (In other words, all magic squares of size p over fields of characteristic p have weight 0 provided $p = 2$ or 3, but in all other cases magic squares of all weights abound.)*

Example.

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

is a magic square of size 5 with nonzero weight in characteristic 5.

Proof. Everything follows if we show that, for $n \geqslant 5$, w is onto and C_n has rank $2n+1$; for once w is onto we have $m_n(F) = z_n(F) + 1$, and rank $(C_n) = 2n+1$ implies $z_n(F) = n^2 - (2n+1) = n(n-2) - 1$. Let D_n denote C_n augmented by a column of 1's. The sum of the first n rows of D_n is $(1, 1, \ldots, 1, n)$ and this is the same as the sum of the next n rows. Thus D_n is row-equivalent to the matrix E_n obtained by replacing its $(n+1)$st row by a row of zeroes. Now E_n has $2n+1$ nonzero rows; we claim E_n has rank $2n+1$. (If so, we are done.)

The first n rows of E_n have their initial nonzero entries in columns $1, 2, \ldots, n$. The $(n+1)$st row is $\underline{0}$, and the following $n-1$ rows of E_n have their initial non-zero entries in columns $n+1, 2n+1, \ldots, (n-1)n+1$. The last two rows of E_n are $(\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_n, 1)$ and $(\underline{e}_n, \underline{e}_{n-1}, \ldots, \underline{e}_1, 1)$. Subtract the first row from the next-to-last to make it $(\underline{0}, \underline{e}_2 - \underline{e}_1, \ldots, \underline{e}_n - \underline{e}_1, 0)$ and subtract the nth row from the last to make it $(\underline{0}, \underline{e}_{n-1} - \underline{e}_n, \ldots, \underline{e}_1 - \underline{e}_n, 0)$. Finally, add the $(n+2)$nd row (the one just under the row of zeroes) to the next-to-last, making it

$$(\underline{0}, \underline{e}_2 - \underline{e}_1 + \underline{j}, \underline{e}_3 - \underline{e}_1, \ldots, \underline{e}_n - \underline{e}_1, 1).$$

We now count to verify that the $2n + 1$ nonzero rows of our reduced version of E_n have their initial nonzero entries in distinct columns. We already have initial nonzero entries in columns 1 through n and $n + 1, 2n + 1, \ldots, (n - 1)n + 1$. Since

$$\underline{e}_2 - \underline{e}_1 + j = (0, 2, 1, 1, \ldots, 1),$$

the next-to-last row has its initial nonzero entry in column $n + 2$ if char. $F \neq 2$ and $n + 3$ if char. $F = 2$. Since $\underline{e}_{n-1} - \underline{e}_n = (0, \ldots, 0, 1, -1)$, the last row has its initial nonzero entry in column $2n - 1$. Finally, since $n \geq 5$ we have $n + 3 < 2n - 1$; thus the $2n + 1$ initial nonzero entries of the rows do indeed occur in distinct columns. Done!

COROLLARY. *When F is infinite, the probability that an $n \times n$ matrix chosen at random will be a magic square is* 0. *When F is finite, with cardinality q, the probability is q^{-3} when $n = 2$, q^{-7} when $n = 4$ and q is even, and q^{-2n} in all other cases.*

REFERENCES

1. R. K. Guy, editor, Reviews in Number Theory 1973–83, American Mathematical Society, 1984.
2. J. E. Ward III, Vector spaces of magic squares, *Math. Mag.*, 53 (1980) 108–111.

The Jerusalem Ticket Problem

JANET CRANDELL TREMAIN

JANET CRANDELL TREMAIN: Currently I am a first year graduate student in mathematics at the University of Missouri in Columbia, Missouri. I've completed a master's degree with a thesis in modal logic. During the last two years I've attended mathematical lectures and conferences in Denmark, Poland, Germany, Switzerland, and Israel. I hope to take some graduate mathematics courses at Princeton and to eventually be a mathematical researcher.

Introduction. As one's mathematical knowledge increases, so does an ability to intuit the mathematically "obvious." This sort of intuition can lead to immeasurable frustration for mathematicians attempting unencumbered lives in a world maintained by nonmathematical clerks. One such case in point is that of Professor Joram Lindenstrauss, mathematical researcher at the Hebrew University of Jerusalem, Israel.

In the spring of 1986, Professor Lindenstrauss approached the ticket counter of Jerusalem Theater to purchase festival tickets for his large family and some visitors.

He arrived on the first day to insure receiving the advertised free ticket per every three purchased from distinct performances. After presenting his large order involving varying numbers of tickets for selections of music, dance, and theater, he requested one-fourth of his order for free as the advertisement had implied.

The ticket agent had no idea he was dealing with a famous mathematician, and since he could not decompose the order into the appropriate groups of three for each free ticket he claimed it would not result in one-fourth the order free.

After recovering from disbelief at what the agent could not grasp, Professor Lindenstrauss sketched out a simple sufficient condition for the "Ticket Problem" (this subcase is subsumed under case one of the theorem below).

The ticket agent was not mathematically inclined and stubbornly stood by his statement that the order could not be legitimately decomposed as he reviewed the unfamiliar figures this customer had presented.

Realizing the difficulty, Professor Lindenstrauss physically grouped his tickets, thereby convincing the chagrined clerk.

In what follows, we will give a complete solution to the "Jerusalem Ticket Problem."

The problem. What is the decomposition of a ticket order maximizing the number of free tickets received if one is received per three purchased for distinct events?

The solution.

Notation

(1) For $1 \le i \le n$, a person buys a_i tickets for performances A_i. We may assume that $a_1 \ge a_2 \ge \cdots \ge a_n$.

(2) Let $l = [\![\sum_{i=1}^{n} a_i/4]\!]$, i.e., an upper bound for the number of free tickets.
(3) $[\![a]\!]$ represents the greatest integer less than or equal to a.
(4) Let l^* = the actual number of free tickets a person will receive given a purchase.
(5) Let $\hat{l} = \max\{m|$ the total tickets ordered can be decomposed into m disjoint sets of three tickets each with no two tickets for the same performance in one set by our counting procedure$\}$.

Note: $\hat{l}$ is given by our counting procedure.

THEOREM. *If one buys a_i tickets for performance A_i, $1 \le i \le n$, and $a_1 \ge a_2 \ge \cdots \ge a_n$, then*:
(1) *If* $\sum_{i=2}^{n} a_i \ge 2a_1$, *then* $l^* = l$.
(2) *If* $\sum_{i=2}^{n} a_i < 2a_1$ *and* $\sum_{i=3}^{n} a_i \le a_2$, *then* $l^* = \min\{\sum_{i=3}^{n} a_i, l\}$.
(3) *If* $\sum_{i=2}^{n} a_i < 2a_1$ *and* $\sum_{i=3}^{n} a_i > a_2$, *then* $l^* = \min\{[\![\sum_{i=2}^{n} a_i/2]\!], l\}$.

Note: The proof of the theorem is deferred.

Remark. Our goal is to determine the actual number of free tickets a person will receive, l^*, given a purchase. It is clear one can get no more free tickets than the total number of tickets divided by four, so we will call this maximum l. By applying a technique for breaking down the order into the greatest possible number of groups of three tickets for distinct events, we will find a way of computing the value of $\hat{l}$. Later, our lemma will give us the actual number of free tickets by the relationship:

$$l^* = \min\{l, \hat{l}\}.$$

LEMMA.

$$l^* = \min\{\hat{l}, l\}.$$

Proof. Clearly, $l^* \le l$ and $l^* \le \hat{l}$, and hence $l^* \le \min\{\hat{l}, l\}$. Also, if $\hat{l} \le l$ then the person is eligible for at most $\hat{l}$ free tickets and we use up $3\hat{l}$ tickets which must be paid for to get the $\hat{l}$ free ones.

So altogether, we use $4\hat{l} \le 4l \le \sum_{i=1}^{n} a_i$ = the total number of tickets. That is, the person is eligible for at most $\hat{l}$ free tickets and there are enough groups of three for distinct events to produce this. Also, there are enough tickets left over for the person to get the $\hat{l}$ free ones. Therefore, in this case, $l^* = \hat{l} = \min\{\hat{l}, l\}$.

Finally, if $l < \hat{l}$, then the person is eligible for $\hat{l}$ free tickets but l is the maximum number possible. So, the person gets l free tickets. That is, in this case we also have that $l^* = l = \min\{\hat{l}, l\}$.

Counting procedure. *Recall*: The a_i represent the cardinality of the A_i and we are assuming that $a_1 \ge a_2 \ge \cdots \ge a_n$.

Now we place the tickets in rows with the tickets for performance A_i in the ith row. We will consider the three cases of the theorem separately.

Case 1. We assume that $\sum_{i=2}^{n} a_i \ge 2a_1$. Let $j = [\![\sum_{i=1}^{n} a_i/3]\!]$ and let r be the remainder in $\sum_{i=1}^{n} a_i/3$, where $r = 0$, 1, or 2. Now form three rows using all of the tickets where row one has length w_1, row two has length w_2, and row three has

length j where,

$$w_1 = \begin{cases} j & \text{if } r = 0 \\ j+1 & \text{if } r = 1 \text{ or } 2 \end{cases}$$

$$w_2 = \begin{cases} j & \text{if } r = 0 \text{ or } 1 \\ j+1 & \text{if } r = 2, \end{cases}$$

using the following procedure:

Line up the A_1 tickets one after another in row one. Now line up the A_2 tickets after the A_1 tickets and go to A_3. Eventually a portion of the tickets from some A_i will finish row one, i.e., we will have w_1 tickets in row one. Start row two with the rest of the tickets from this A_i and continue until row two has w_2 tickets and row three has just j tickets.

Now we divide the tickets into groups of three by grouping the three tickets in column one together, the three tickets in column two together, until we reach column j. It is clear that since $a_1 \geq a_2 \geq \cdots \geq a_n$, our groups will contain tickets from distinct performances and all the tickets will be grouped except the last $r = 0$, 1, or 2. (See FIGURES 1 and 2.)

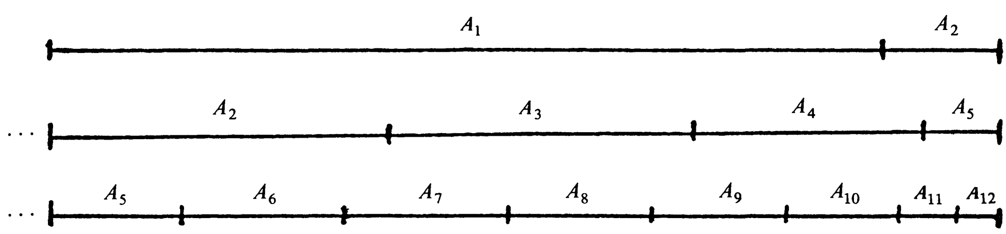


FIG. 1. Arrangement of tickets in Case 1.

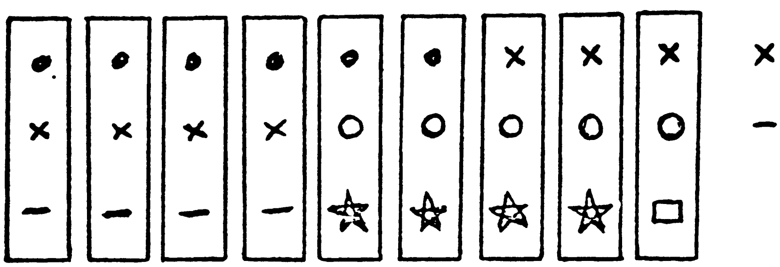


FIG. 2. Grouping tickets in Case 1.

Case 2. We assume that $\Sigma_{i=2}^{n} a_i < 2a_1$ and $\Sigma_{i=3}^{n} a_i \leq a_2$.

Form row one with A_1 tickets. Form rows two and three by placing the A_2 tickets in row two and then placing the remaining tickets in row three. (See FIGURE 3.)

Now we group the tickets into groups of three for distinct performances as in case one. Clearly we have exactly as many groups as $\Sigma_{i=3}^{n} a_i$.

Case 3. We assume that $\Sigma_{i=2}^{n} a_i < 2a_1$ and $\Sigma_{i=3}^{n} a_i > a_2$.

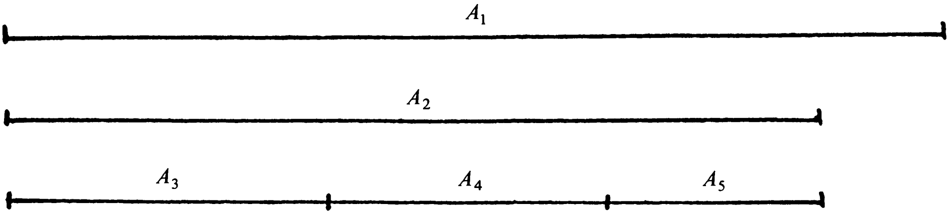


FIG. 3. Arrangement of tickets in Case 2.

Form row one with the tickets A_1. Form rows two and three by placing $k = [\![\sum_{i=2}^{n} a_i/2]\!]$ terms of $A_2, A_3, \ldots, A_n$ consecutively in row two if n is even ($k+1$ if n is odd), and then placing the remaining k tickets in row three.

After grouping as in case one we get precisely k groups of three distinct tickets each. (See FIGURE 4.)

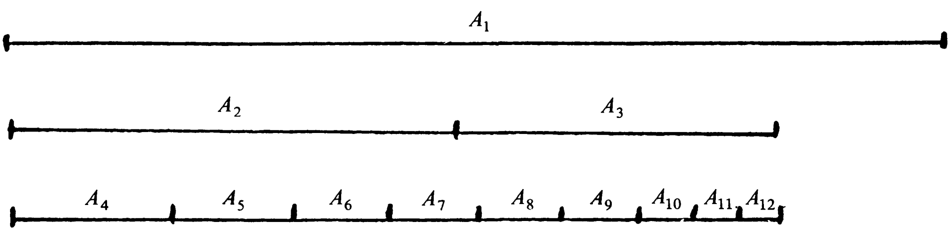


FIG. 4. Arrangement of tickets in Case 3.

Note: The counting procedure gives us $\hat{l}$.
Now we are ready to prove our theorem.

Proof.

Case 1. Let $j = [\![\sum_{i=1}^{n} a_i/3]\!]$ and let r be the remainder in $\sum_{i=1}^{n} a_i/3$ where $r = 0, 1, 2$. Now we apply the counting procedure. As we saw earlier, this procedure divides all the tickets except $r = 1$, or 2 into groups of three distinct tickets. Therefore, $\hat{l} = [\![\sum_{i=1}^{n} a_i/3]\!]$ since we have grouped all tickets except for one or two.

Hence,

$$\hat{l} = \left[\!\!\left[\sum_{i=1}^{n} a_i/3 \right]\!\!\right] \geq \left[\!\!\left[\sum_{i=1}^{n} a_i/4 \right]\!\!\right] = l,$$

and so $l^* = \min\{\hat{l}, l\} = l$. This completes the proof of case 1.

Case 2. It suffices to prove that $\sum_{i=3}^{n} a_i = \hat{l}$, by our lemma. If we divide the tickets into groups of three different kinds per group, it follows that each group contains at least one ticket from one of the performances A_i for $3 \leq i \leq n$. Hence, $\hat{l} \leq \sum_{i=3}^{n} a_i$.

But, lining up performances A_3 through A_n end to end as row three (and since $\sum_{i=3}^{n} a_i \leq a_2 \leq a_1$), by the counting procedure: $\sum_{i=3}^{n} a_i \leq \hat{l}$. Therefore, $\hat{l} = \sum_{i=3}^{n} a_i$. This completes the proof of case 2.

Case 3. Let $k = \llbracket \sum_{i=2}^{n} a_i/2 \rrbracket$ and note that $\sum_{i=3}^{n} a_i > a_2$ implies $k \geq a_2$. After applying the counting procedure we get $\hat{l} \geq k$. But, any division of the tickets into groups of three different tickets has the property that each group contains at least two tickets from the performances $A_2, \ldots, A_n$. Hence, the number of groups is less than or equal to $\llbracket \sum_{i=2}^{n} a_i/2 \rrbracket = k$. That is, $\hat{l} \leq k$. Therefore, $\hat{l} = k$. But now, $l^* = \min\{l, \hat{l}\} = \min\{l, k\} = \min\{l, \llbracket \sum_{i=2}^{n} a_i/2 \rrbracket$. This completes the proof of case three and hence of the theorem. An examination of the proof of the theorem gives the following corollary.

COROLLARY. *The counting procedure always gives $\hat{l}$. Furthermore, $\hat{l}$ will be the correct number of free tickets a person is eligible for and we can add j free tickets of any kind where $j = \hat{l} - l$ when $\hat{l} \geq l$.*

If $\hat{l} \leq l$ then $\hat{l}$ is the number of free tickets the person will actually receive. In other words, these two processes give us a decomposition method yielding the optimal grouping.

Variations. In addition to the obvious variations of this problem (e.g., receiving two free tickets for each four purchased), an interesting variation occurs by excluding certain of the expensive tickets from qualifying as free.

Another even more interesting variation occurs by assuming the tickets have different prices and that we want to pay the least possible. Note that in this case, we may need to take fewer than the optimal number of free ones as the following example illustrates.

Let $a_1 = 3$ tickets at \$1/each, $a_2 = 3$ tickets at \$1/each, $a_3 = 1$ ticket at \$10 each, and $a_4 = 1$ ticket at \$10/each. This falls into case two of our theorem. Our previous rearrangement without regards to cost can be seen in FIGURE 5.

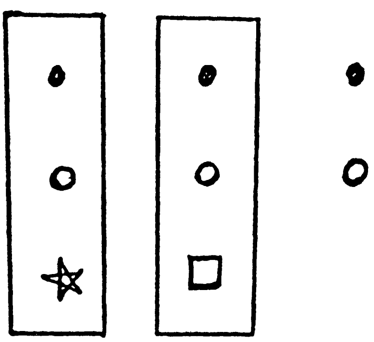


FIG. 5. Old grouping.

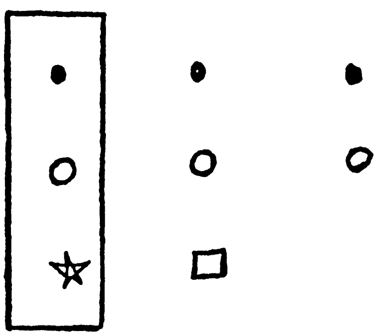


FIG. 6. Cost optimizing grouping.

However, to optimize cost this would not be the correct choice since we would pay \$24 for tickets in order to receive \$2 worth free whereas we could pay \$16 for tickets in order to get \$10 worth free if we make the choice where now a \$10 ticket is free as in FIGURE 6.

LETTERS TO THE EDITOR

Editor:

Regarding Barry Cipra's "Introduction to the Ising Model" [**3**], we find fault with neither its prose nor its derivations, which could set a new standard for tutorial articles in applied mathematics. Yet, as the author's stated intent was to attract attention to the model, his concluding paragraphs might have offered it as a springboard to understanding the more complicated spin glass models which have been used in the analysis of "neural networks." The explosive growth of international and interdisciplinary fascination with these connectionist models is surely one of the cultural phenomena of science in this decade [**4**]. A spark was provided by Hopfield [**5**], whose expression

$$H(\mathbf{S}) = -\sum_{j=i+1}^{N}\sum_{i=1}^{N-1} S_i S_j T_{ij} - \sum_{i=1}^{N} S_i U_i \tag{1}$$

for the "computational energy" of the network is analogous to Cipra's equation (1.1). Here S_i is the state of the ith neuron-like element—either firing at its peak rate ($S_i = 1$) or resting ($S_i = 0$); and the connectivity matrix $\|T_{ij}\|$ gives the strength of the "synapse" through which the ith unit excites (if $T_{ij} > 0$) or inhibits (if $T_{ij} < 0$) the jth unit. The state of the ith unit is decided by the threshold test applied to its input,

$$X_i = \sum_{j=1}^{N} S_j T_{ji} + U_i. \tag{2.1}$$

The binary "McCulloch-Pitts neuron" obeys the rule

$$S_i = \begin{cases} 1 & \text{if } X_i > 0 \\ 0 & \text{if } X_i \le 0. \end{cases} \tag{2.2}$$

When the T-matrix is real-valued and symmetric, with all zeros on the diagonal, the network evolves toward stable states $\mathbf{s}$ which correspond to local minima of the computational energy. The "energy landscape" can be configured so that these local minima correspond to solutions of constrained optimization and pattern recognition problems [**9**, **10**]. In the latter case, the vector $\mathbf{U}$ of inputs to the N units might represent the pixel pattern on a retina.

A provocative paper by Ackley, Hinton and Sejnowski [**1**] proposed simulated annealing to dislodge the Hopfield network from local minima and enable it to settle into states of still lower energy which would represent better (if still suboptimal) solutions. The network is "heated" by the addition of noise to the input of each unit. When these noises are independent, identically distributed random variables, the state $\mathbf{S}$ takes a random walk on the 2^N vertices of a hypercube. The stationary distribution,

$$Pr(\mathbf{S} = \mathbf{s}) = \exp[-\beta H(\mathbf{s})] \Big/ \sum_{\mathbf{s}'} \exp[-\beta H(\mathbf{s}')], \tag{3}$$

is formally equivalent to Cipra's equation (1.3). The assertion of Ackley, Hinton and Sejnowski, that $T = 1/\beta$ is the root mean intensity of noise described by a logistic distribution, is not powerfully motivated. Shaw et al. [**8**] had earlier arrived at an

expression like (3) in which β is a "smearing factor" determined from details of a stochastic model of the chemical synapse.

It was over a hundred years ago that Gibbs sought time-invariant solutions to a Liouville equation in which the independent variables were the Hamiltonian coordinates of a multiparticle system and the dependent variable was the probability of the system being in a given state. He arrived at a canonical ensemble in which "the index of probability [i.e., the log-probability] is a linear function of the energy" of the state. This result is expressed by equation (3), usually called a Boltzmann distribution. Other functions of the energy, however, will serve this purpose; and the fact that the linear dependence (of log-probability on energy) maximizes the entropy of the system is not necessarily germane to the question. Belief in the possibility of a mathematical treatment of biological intelligence, patterned after statistical thermodynamics, goes back at least as far as the works of John von Neumann (published posthumously). For this belief to find expression in contemporary neural network research is not surprising. The mathematician who studies this work must be slightly bewildered by derivations which appeal to analogies with statistical physics, some of which are complicated by psychological theory [7]. The validity of the Boltzmann distribution, in the context of the connectionist paradigm, is solely dependent on the existence of models which give rise to it. As far as real neuronal networks are concerned, the laboratory experiments which would verify the result have yet to be defined.

The computational technique of simulated annealing traces its roots to the Metropolis [6] algorithm, which updates the state of an N-particle system according to a stochastic model in which the Boltzmann distribution is expressly assumed beforehand. Our own research has led to an alternative derivation of C. R. Darnafalski, the amateur mathematician whose unpublished essays we have cited elsewhere [2]. The stochastic model, adapted to the present problem, leads to a simple algorithm as follows. Pick an integer $i \in \{1, \dots, N\}$ at random. Compute X_i according to (2.1). Modify X_i by the addition of a real random variable, call it Y_1, which is symmetrically distributed about a mean of zero. Compute S_i according to (2.2). These steps are iterated indefinitely with independent, identically distributed random numbers $\{Y_k, k = 1, 2, \dots\}$. It is not hard to see that this gives rise to a sequence $\{\mathbf{S}_k, k = 1, 2, \dots\}$ of states which constitute a Markov chain. Nonzero probabilities can be attributed to transitions which involve at most one component of the state vector. With no external input ($\mathbf{U} = \mathbf{0}$), these probabilities depend on the T_{ij} and the distribution of Y. When Hopfield's conditions are obeyed by the former, the stationary distribution can be derived analytically. This distribution is

$$\Pr(\mathbf{S} = \mathbf{s}) = Z^{-1}\exp\left\{ \sum_{j=i+1}^{N} \sum_{i=1}^{N-1} s_i s_j \log\left\{F(T_{ij})/\left[1 - F(T_{ij})\right]\right\}\right\},$$

in which F is the (cumulative) distribution function of Y and the denominator Z is the sum over all states which normalizes the discrete density. The assumption of logistic noise, as

$$F(y) = 1/(1 + e^{-\beta y}), \quad -\infty < y < \infty,$$

gives the last equation a particularly simple form (3).

The central theme of Cipra's article is the discovery of closed-form expressions for the partition function Z. These would allow the analyst to identify the critical

temperatures of phase transitions in the physical systems described by spin glass models, of which Ising's is the progenitor. At the present time, however, the creation of computational devices to accelerate the simulation of such systems is becoming a major R & D initiative, motivated by the quest for computers that can carry out associative mappings and solve optimization problems using highly parallel, simple computing elements. The same phenomenon which the physicist calls a phase transition is, in the connectionist view, the conclusion of a state space search that constitutes an intelligent response to given stimuli. Some will argue the futility of the quest for such "neutral architectures"; but it seems not unlikely that the effort will produce tangible products. The best insights of some talented mathematicians will be required to make sense of what these products can and cannot do.

James P. Coughlin
Towson State University
Towson, Maryland

Robert H. Baran
Naval Surface Warfare Center
Silver Spring, MD

REFERENCES

1. D. H. Ackley et al., A learning algorithm for Boltzmann machines, *Cognitive Science*, 9 (1985) 147–169.
2. R. H. Baran and J. P. Coughlin, Comments on a population model of China, *Automatica*, 22 (1986) 255–256.
3. B. Cipra, An introduction to the Ising model, *American Mathematical Monthly*, 94 (1987) 937–959.
4. R. Hecht-Nielson, Neurocomputing: picking the human brain, *IEEE Spectrum* (March, 1988).
5. J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, *Proc. Nat. Acad. Sci. USA*, 79 (1982) 2554–2558.
6. N. Metropolis et al., Equation of state calculations by fast computing machines, *J. Chem. Phys.*, 21 (1953) 1087–1092.
7. D. E. Rumelhart et al., Parallel Distributed Processing: Exploration in the Microstructure of Cognition, vol. 1, MIT Press, Cambridge, MA, 1986.
8. D. L. Shaw and K. J. Roney, Analytic solution of a neural network theory based on an Ising spin system analogy, *Physics Letters*, 74A (1979) 146–149.
9. D. W. Tank and J. J. Hopfield, Simple "neural" optimization networks, *IEEE Trans. Circuits and Systems*, 33 (1986) 533–541.
10 ______, Collective computation in neuronlike networks, *Scientific American* (December 1987) 104–114.

NOTES

EDITED BY DAVID J. HALLENBECK, DENNIS DETURCK, AND ANITA E. SOLOW

Decomposition of Bipartite Graphs Into Paths

KIRAN B. CHILAKAMARRI
Department of Aeronautical and Astronautical Engineering, Ohio State University, Columbus, Ohio 43210

Liu and Zaks prove in [**1**], among other results, that bipartite graph $K_{n,n+1}$ can be decomposed into paths $P_2, P_4, \ldots, P_{2n}$ of length (i.e., number of edges in the path) $2, 4, \ldots, 2n$, respectively, for all odd n. They conjecture that this cannot be done for any even integer n. This conjecture is obviously prompted by observing that from $K_{2,3}$ one cannot obtain both P_2 and P_4 at the same time. The purpose of this note is to show that we can indeed decompose $K_{n,n+1}$ into paths $P_2, P_4, \ldots, P_{2n}$ for all even $n \geq 4$ and failure to do so occurs only for $n = 2$.

We need a few observations to obtain the result. First consider an $n \times m$ array $R_{n,m}$ of points where a point in the ith row and jth column is identified with edge (x_i, y_j) of a bipartite graph $K_{n,m}$ with vertex set $\{x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_m\}$ and edges joining any x_p to any y_q for $1 \leq p \leq n$ and $1 \leq q \leq m$. Any path on the points of $R_{n,m}$ with the properties: (1) travels only along rows or columns (2) uses at most two points from any row or column and (3) whose end points do not lie in the same row or column, defines a unique path in $K_{n,m}$. If a path with (1), (2), and (3) in $R_{n,m}$ uses N points then the corresponding path in $K_{n,m}$ uses exactly N edges and has no repeated vertices.

Now the problem of decomposing $K_{n,n+1}$ into paths $P_2, P_4, \ldots, P_{2n}$ for even $n \geq 4$, reduces to covering $R_{n,n+1}$ with paths using $2, 4, \ldots, 2n$ points and each satisfying conditions (1), (2), and (3). We will do this for $n = 4, 6$, and 8. For an even integer n beyond 8 the construction we give can be continued. A general procedure for the construction can be given but we will not do so since it is clumsy and only obscures the procedure.

One can read off paths in $K_{4,5}$ from FIG. 1. For example (x_1, y_1) and (x_3, y_1) form P_2 in $K_{4,5}$ Similarly (x_2, y_1), (x_2, y_2), (x_1, y_2), (x_1, y_3), (x_4, y_3), and (x_4, y_5) form P_6 in $K_{4,5}$. For illustration purposes we show the decomposition of $K_{4,5}$ in FIG. 2.

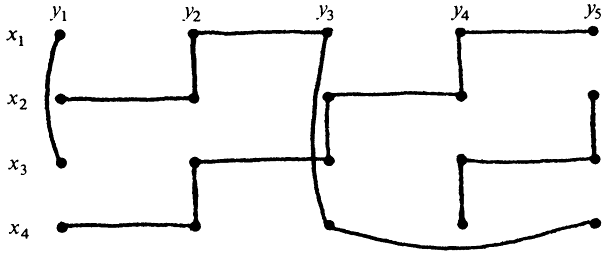


FIG. 1. $n = 4$. Decomposition of $K_{4,5}$ into P_2, P_4, P_6, P_8.

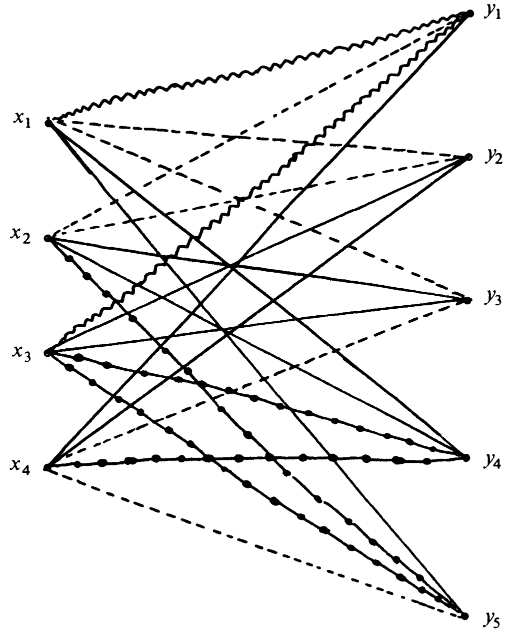


FIG. 2.

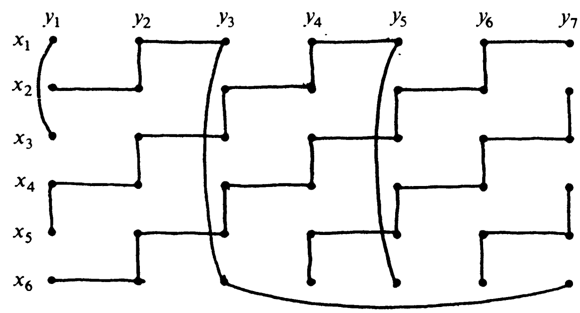


FIG. 3. $n = 6$. Decomposition of $K_{6,7}$ into P_2, P_4, P_6, P_8, P_{10}, P_{12}.

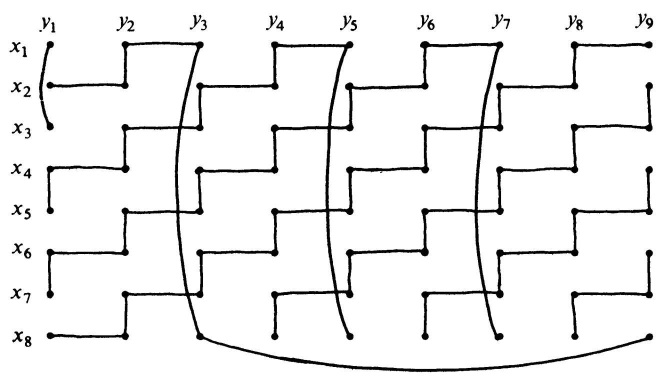


FIG. 4. $n = 8$. Decomposition of $K_{8,9}$ into $P_2, P_4, \ldots, P_{16}$.

The author is grateful to Professor S. Zaks for suggesting the problem.

REFERENCES

1. S. Zaks and C. L. Liu, Decomposition of Graphs into Trees, Proc. 8th S. E. conf., *Combinatorics, Graph Theory and Computing*, 643–654.

Fermat's Last Theorem in Combinatorial Form

W. V. QUINE
Department of Philosophy, Harvard University, Cambridge, MA 02138

Fermat's Last Theorem can be vividly stated in terms of sorting objects into a row of bins, some of which are red, some blue, and the rest unpainted. The theorem amounts to saying that when there are more than two objects, the following statement is never true:

STATEMENT. *The number of ways of sorting them that shun both colors is equal to the number of ways that shun neither.*

I shall show that this statement is equivalent to Fermat's equation $x^n + y^n = z^n$, when n is the number of objects, z is the number of bins, x is the number of bins that are not red, and y is the number of bins that are not blue. There are z^n ways of sorting the objects into bins; x^n of these ways shun red and y^n of them shun blue. So where A is the number of ways that shun both colors, B is the number of ways that shun red but not blue, C is the number of ways that shun blue but not red, and D is the number of ways that shun neither, we have

$$x^n = A + B, \qquad y^n = A + C, \qquad z^n = A + B + C + D.$$

These equations give $x^n + y^n = z^n$ if and only if $A = D$, which is to say, if and only if the statement above holds.

A Difference Equation for Strings of Ones

ANDREW SIMOSON
Mathematics Department, King College, Bristol, TN 37620

The classic proof [**2**, pp. 122–3] that

$$\text{almost every number in its binary expansion contains arbitrarily long strings of the digit 1} \qquad (*)$$

can be translated into an elegant difference equation argument, as shown below.

In this note, for each number such as $\frac{1}{2}$ which has a dual binary expansion, interpret its binary expansion as the one terminating in all zeroes. Let I be the half open interval $[0, 1)$. For each positive integer n, let B_n be the set of numbers in I containing no string of n 1's in their binary expansions. To prove $(*)$ it is sufficient to show that B_n has measure zero. For all integers $k \geq 0$, $0 \leq j < n$, let $Y_j(k)$ be the

set of all numbers in I such that the first k digits of their binary expansion

i. terminate with a string of exactly j 1's;
ii. contain no string of n 1's.

Note that $Y_j(k)$ is the union of pairwise disjoint intervals of length $1/2^k$ for $j \leq k$, and that $Y_j(k)$ is empty for $j > k$. For example, if $n = 2$ then $Y_0(3) = [0, 1/8) \cup [1/4, 3/8) \cup [1/2, 5/8)$ and $Y_1(3) = [1/8, 1/4) \cup [5/8, 3/4)$. Also note that

$$B_n \subset \bigcup_{j=0}^{n-1} Y_j(k) \quad \text{for all } k \geq 0. \tag{$**$}$$

Note that for $0 \leq j < n - 1$, "half" of the numbers that are in $Y_j(k)$ are those numbers that comprise the set $Y_{j+1}(k+1)$, whereas "half" of the numbers in $Y_{n-1}(k)$ have a string of n 1's in the first $k+1$ digits of their binary expansions. Furthermore for $0 \leq j < n$, the other "halves" of $Y_j(k)$ make up $Y_0(k+1)$. Hence, if $y_j(k)$ denotes the measure of $Y_j(k)$, then $y_0(k+1) = \frac{1}{2}\sum_{j=0}^{n-1} y_j(k)$ and $y_{j+1}(k+1) = \frac{1}{2}y_j(k)$ for all $k \geq 0$, $0 \leq j < n - 1$. Let $y(k)$ be the transpose of the n-tuple $(y_0(k), y_1(k), \ldots, y_{n-1}(k))$. These relations define the difference equation

$$y(k+1) = Ay(k) \text{ for all } k \geq 0, \text{ where } A \text{ and } y(0) \text{ are the } n \times n \text{ matrix} \tag{$***$}$$

and $n \times 1$ vector:

$$A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & & & & \cdots & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & & & \cdots & 0 \\ 0 & \frac{1}{2} & 0 & 0 & & \cdots & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & \cdots & 0 \\ \vdots & & & & \ddots & & \vdots \\ 0 & & & & 0 & 0 \ \ \frac{1}{2} & 0 \end{bmatrix} \quad \text{and } y(0) = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Then for all $k \geq 0$, by $(**)$ and the definition of $y(k)$ it follows that

$$\begin{aligned}(\text{the measure of } B_n) &\leq (\text{the sum of the components of } y(k)) \\ &= (\text{the sum of the components of } A^k y(0)).\end{aligned}$$

At this point, recall some linear algebra:

(1) If all nonzero entries of an $n \times n$ matrix A are positive then the dominant eigenvalue of A is a real number between 0 and the maximum of c_j, $1 \leq j \leq n$, where c_j is the sum of the components of column j of A, [**3**, pp. 193–5].

(2) If all the eigenvalues of an $n \times n$ matrix A have magnitude less than 1 then A^k converges to the zero matrix as k increases, [**3**, p. 157].

By (1) the dominant eigenvalue of A is a positive number less than or equal to 1. But 1 is not an eigenvalue of A. To see this note that the matrix $2A$ of $(***)$ is the companion matrix of the polynomial equation $1 + x + x^2 + \cdots + x^{n-1} - x^n = 0$, (which is also the characteristic equation of 2A). See [**4**, p. 25], for example. This

polynomial equation (for $n > 1$) can be written as $(1 - 2x^n + x^{n+1})/(1 - x) = 0$. Since 2 is not a root of this equation and since the eigenvalues of A are $\frac{1}{2}$ the roots of this equation, then 1 is not an eigenvalue of A. Therefore, by (2), $A^k y(0)$ converges to the zero vector as k increases, and so B_n has measure zero.

The above argument generalizes for the following extension of $(*)$: almost every number in its binary expansion contains any given binary string. To outline this procedure, let $S = a_1 a_2 \cdots a_n$ be a string of binary digits of length n. Let B_n be those numbers in I containing no instance of S in their binary expansions. For all integers $k \geq 0$, $0 \leq j < n$, let $Y_j(k)$ be the set of all numbers in I such that the first k digits of their binary expansion

i. terminate in the string $a_1 a_2 \cdots a_j$, but in no longer prefix of S;
ii. contain no instance of S.

Define $y(k)$ as before, yielding the difference equation $y(k+1) = Ay(k)$. This matrix A resembles that of $(***)$ in that

i. each of its $2n - 1$ nonzero entries is $\frac{1}{2}$;
ii. each of its first $n - 1$ columns contain exactly 2 nonzero terms;
iii. exactly n of the nonzero entries are in the upper triangular region of A;
iv. the remaining $n - 1$ entries of A are on the $(j+1, j)$ diagonal of A, where $1 \leqslant j \leqslant n$.

For example, if S is the string 1101, then the matrix A is

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 \end{bmatrix}.$$

For matrices such as these it follows that 1 is not an eigenvalue. To see this let e be an eigenvector for A associated with a supposed eigenvalue of 1. Comparing the sum of the components of Ae with the sum of the components of e forces the last component of e to be 0, which upon successive back substitutions force the remaining components of e to be 0, which means that 1 is not an eigenvalue of A. The remainder of the argument is as before.

Furthermore, the above argument is valid for the analogous result in base ten (or any other base); for example to show the analog of $(*)$ in base ten, in the matrix A of $(***)$ simply replace the first row of $\frac{1}{2}$'s with 9/10's and replace the remaining $(\frac{1}{2})$'s with 1/10's.

For a contrasting, more measure theoretic approach to this same material see [**1**, p. 43–8].

REFERENCES

1. M. Adams and V. Guillemin, Measure Theory and Probability, Wadsworth & Brooks/Cole, Monterey, CA, 1986.
2. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, Oxford-Clarendon Press, London, 1960.
3. D. G. Luenberger, Introduction to Dynamic Systems, John Wiley & Sons, New York, 1979.
4. H. Wilf, Mathematics for the Physical Sciences, Dover Publications, New York, 1978.

Continued Fractions and Differentiability of Functions

ALEC NORTON
Department of Mathematics, Boston University, Boston, MA 02215

The "ruler function" $f\colon \mathbb{R} \to \mathbb{R}$ is defined by the rule

$$f(x) = \begin{cases} 0 & \text{if } x \text{ is irrational} \\ 1/q & \text{if } x = p/q \text{ in lowest terms (with } q > 0). \end{cases}$$

(By convention, if x is an integer, then $x = x/1$ in lowest terms, e.g., $f(0) = 1$.) This is a well-known example of a function continuous at the irrationals but discontinuous at the rationals. Suppose that we wish to improve continuity at the irrationals to differentiability. The ruler function is differentiable nowhere, but a simple modification will make it differentiable almost everywhere.

The idea of modifying f to produce points of differentiability has been around for some time, as has the more general question of differentiability in the face of a dense set of discontinuities, e.g., **[1]**, **[2]**, **[3]**, **[5]**, **[6]**, **[8]**. However, the literature is fragmentary and not widely known. Furthermore, the point of view of Diophantine approximation taken here has not been fully exploited, nor, apparently, have derivatives higher than the first been discussed in this context. To do so is to provide a beautiful connection between the notions (1) degree of differentiability and (2) degree of approximability by points of a dense set (such as the rationals).

First we need a definition. Let $k \geq 0$ be an integer. A function $f\colon \mathbb{R} \to \mathbb{R}$ is k times (Peano) differentiable at a point x if there is a polynomial $P_x(t)$ of degree $\leq k$ such that $P_x(0) = f(x)$ and

$$\frac{f(x+t) - P_x(t)}{t^k} \to 0 \quad \text{as } t \to 0.$$

(Note that when $k = 0$ this is just the definition of continuity.) Here P_x is the kth order Taylor polynomial for f at x. If f is k times differentiable at x, define the kth (Peano) derivative of f at x by $f^{(k)}(x) = (\text{coefficient of } t^k \text{ in } P_x(t)) \cdot k!$. (It should be remarked that for $k > 1$ this definition is slightly more general than the usual one, which requires that $f^{(k-1)}$ exist in a neighborhood of x for $f^{(k)}(x)$ to be defined. However, the Peano derivative agrees with the ordinary higher derivative whenever the latter is defined, and has the virtue of allowing us to discuss higher derivatives in the context of a dense set of discontinuities. In this context, we will omit the word "Peano" from now on without danger of ambiguity.) A function f is infinitely differentiable at x if it is k times differentiable at x for all k.

Define $g\colon \mathbb{R} \to \mathbb{R}$ by

$$g(x) = \begin{cases} 0 & \text{if } x \text{ is irrational} \\ 1/2^q & \text{if } x = p/q \text{ in lowest terms (with } q > 0). \end{cases}$$

PROPOSITION. *There is a partition of the irrationals $\mathbb{R} \setminus \mathbb{Q}$ into uncountable dense sets $\Theta_0, \Theta_1, \ldots, \Theta_\infty$ with the property that g is infinitely differentiable at each point of Θ_∞, and for each $k \geq 0$, if $\theta \in \Theta_k$, then g is k times differentiable but not $(k+1)$ times differentiable at θ. Furthermore, Θ_0 is residual in $\mathbb{R}$ (i.e., contains a countable intersection of open dense sets), while $\mathbb{R} \setminus \Theta_\infty$ has Lebesgue measure zero.*

Proof. The proof uses the ideas of Diophantine approximation (good references are [4] or [7], among many others); in fact the partition $\{\Theta_k\}$ will merely be a partition according to the rate of divergence of the coefficients in the continued fraction expansions of the irrationals (see below).

We henceforth use θ to denote irrationals. First note that, for any k, if the kth derivative $g^{(k)}(\theta)$ exists, it must be zero. We can see this by induction. For $k = 1$, this follows because $g(\theta) = 0$ and $g \geq 0$. Suppose $g'(\theta) = \cdots = g^{(k-1)}(\theta) = 0$, and g is k times differentiable at θ. Then

$$\frac{g(\theta + t) - P_\theta(t)}{t^k} \to 0 \quad \text{as } t \to 0,$$

so

$$\frac{1/2^q - g^{(k)}(\theta)(\theta - p/q)^k/k!}{(\theta - p/q)^k} \to 0 \quad \text{as } p/q \to \theta.$$

But the quantity $(1/2^q)/(\theta - p/q)^k$ has a limit only if that limit is zero: this is because there will always be rationals with very large denominators that are only moderately near θ. Hence $g^{(k)}(\theta) = 0$.

Since g is periodic with period one, it is enough to consider just the interval $[0,1]$. Each irrational $\theta \in [0,1]$ has a unique infinite continued fraction expansion

$$\theta = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots}}},$$

denoted $[a_1, a_2, a_3, \ldots]$, where the coefficients a_i (called partial quotients) are positive integers. Truncating at the nth stage yields the nth rational convergent $p_n/q_n = [a_1, a_2, \ldots, a_n]$, which has the property that, among rationals p/q with $q \leq q_n$, $|\theta - p/q|$ is minimized when $p/q = p_n/q_n$. It turns out that the sequence $\{q_n\}$ satisfies the relations

$$q_0 = 1, \qquad q_1 = a_1, \qquad q_{n+1} = a_{n+1}q_n + q_{n-1}. \tag{1}$$

Moreover, we have the standard inequalities (which, e.g., follow from eq. (37), p. 237 of [7]):

$$\frac{1}{(2 + a_{n+1})q_n^2} < |\theta - p_n/q_n| < \frac{1}{a_{n+1}q_n^2}. \tag{2}$$

The question of the differentiability of g at θ will hinge on the rate at which the sequence $\{a_i\}$ is increasing, or more exactly, the rate at which the monotone sequence $\{A_n\}$ is increasing, where $A_n = \max\{a_i: i = 1, \ldots, n+1\}$.

For convenience, define a function F: $\mathbb{Z}^+ \to \mathbb{Z}^+$ (depending only on θ) by $F(q_j) = 2 + A_j$ for each j, and $F(q) = 2$ for all q not in the sequence $\{q_j\}$. We will need the following general fact:

$$\text{for all rationals } p/q,\ |\theta - p/q| \geq 1/q^2F(q). \tag{3}$$

To see this, suppose otherwise for some p/q. Then in particular $|\theta - p/q| < 1/2q^2$.

But this implies (e.g. [7, p. 237] that p/q is actually one of the convergents p_n/q_n. Hence $|\theta - p_n/q_n| < 1/q_n^2(2 + A_n)$, but this contradicts (2).

Now suppose g is k times differentiable at θ. Since the first k derivatives of g must therefore be zero, this is true if and only if

$$(1/2^q)\Big/|\theta - p/q|^k \to 0 \quad \text{as } p/q \to \theta$$

By (3), this will be true if $(q^2F(q))^k/2^q \to 0$ as $q \to \infty$, which holds if and only if

$$\left(q_n^2A_n\right)^k/2^{q_n} \to 0 \quad \text{as } n \to \infty.$$

Conversely, suppose $(q_n^2A_n)^k/2^{q_n}$ does not converge to zero. Then for some $\varepsilon > 0$, $\mathbf{A} = \{n\colon (q_n^2A_n)^k/2^{q_n} > \varepsilon\}$ is an infinite set. Furthermore, $\mathbf{A}$ contains infinitely many n such that $A_n = a_{n+1}$. (To see this, choose N so large that $q_n^{2k}/2^{q_n}$ is monotone decreasing for $n > N$. Choose $m > N$ so that $A_m = a_{m+1}$ (possible since $\{A_n\}$ cannot be bounded if $\mathbf{A}$ is infinite). Now pick $p \in \mathbf{A}$ such that $p > m$. If $A_p \neq a_{p+1}$, then $A_p = A_{p-1}$ and $p - 1 \in \mathbf{A}$. Repeating with $p - 1$, we must eventually find $n \in \mathbf{A}$ with $n \geq m$ and $A_n = a_{n+1}$. This can be done for arbitrarily large m.) We have now shown that $(q_n^2a_{n+1})^k/2^{q_n} > \varepsilon$ for infinitely many n. Therefore,

$$(1/2^{q_n})/|\theta - p_n/q_n|^k > \left(q_n^2a_{n+1}\right)^k/2^{q_n} > \varepsilon$$

for infinitely many n (using (2)). Hence g cannot be k-times differentiable at θ.

We have proved that g is k-times differentiable at θ if and only if

$$\left(q_n^2A_n\right)^k\Big/2^{q_n} \to 0 \quad \text{as } n \to \infty. \tag{4}$$

Now we simply define Θ_k to be the set of irrationals θ for which (4) holds for k but not for $k + 1$.

For $k > 0$, the following argument shows that Θ_k is dense and uncountable. Fix k. We inductively define a sequence $\{a_k\}$ of partial quotients as follows. Let $a_1 = 1$, and if $a_1, \ldots, a_n$ have been defined (and hence $q_1, \ldots, q_n$ by means of (1)), let a_{n+1} be the least integer greater than or equal to $q_n^{-3}2^{(q_n/k)}$. Then

$$\begin{aligned}\left(a_{n+1}q_n^2\right)^k/2^{q_n} &\leq \left[\left(q_n^{-3}2^{q_n/k} + 1\right)q_n^2\right]^k\Big/2^{q_n},\\ &= q_n^{-k}2^{q_n}\left[1 + q_n^32^{-q_n/k}\right]^k\Big/2^{q_n}\\ &= q_n^{-k}\left[1 + q_n^32^{-q_n/k}\right]^k \to 0 \quad \text{as } n \to \infty.\end{aligned}$$

On the other hand,

$$\begin{aligned}\left(a_{n+1}q_n^2\right)^{k+1}/2^{q_n} &\geq \left(q_n^{-1}2^{q_n/k}\right)^{k+1}\Big/2^{q_n}\\ &= q_n^{-(k+1)}2^{q_n/k} \to \infty \quad \text{as } n \to \infty.\end{aligned}$$

Since $\{a_i\}$ is eventually increasing, $a_{n+1} = A_n$ for n large enough. This means that $[a_1, a_2, \ldots] \in \Theta_k$ by definition of Θ_k. Indeed, the set $\{[c_1, c_2, c_3, \ldots]\colon c_i = a_i$ or $1 + a_i$ for each $i\}$ is contained in Θ_k by similar reasoning, and is uncountable by virtue of the uniqueness of continued fraction expansions [7]. Furthermore, for any finite sequence $\{b_1, b_2, \ldots, b_m\}$, $[b_1, \ldots, b_m, a_{m+1}, a_{m+2}, \ldots] \in \Theta_k$ since a finite

number of changes in the sequence $\{a_i\}$ cannot affect the limit (4). The reader can readily check that the set of all such points is dense in $[0,1]$, and hence, so is Θ_k.

Define Θ_∞ to be the set of θ such that (4) holds for all $k \in Z^+$. In case $A_n \leq Cq_n^\beta$ for some $\beta, C > 0$, then the corresponding irrational is called *Diophantine*. Any Diophantine irrational is contained in Θ_∞ since for each k, $q_n^{(2+\beta)k}/2^{q_n} \to 0$ as $n \to \infty$. It is well known **[9**, p. 8] that the Diophantine irrationals form a set of measure one in $[0,1]$. Hence Θ_∞ is also of full measure.

On the other hand, the following argument shows that Θ_0 is residual. Let $W(p,q) = \{x \in [0,1]: \ |x - p/q| < 1/2^q\}$. This set is open, and $V_n = \bigcup_{q>n}\bigcup_p W(p,q)$ contains all the rationals, and, hence, is open dense. Let $V = \bigcap_n V_n$. If $\theta \in V$ is irrational, then for infinitely many q, $|\theta - p/q| < 1/2^q < 1/2q^2$ for some p. But, as before, such a q must in fact be a q_n, so by (2) we have

$$\frac{1}{(2 + a_{n+1})q_n^2} < |\theta - p/q| < \frac{1}{\left(q_n^{-2}2^{q_n}\right)q_n^2}.$$

Therefore, $2 + a_{n+1} > q_n^{-2}2^{q_n}$ for infinitely many n, which implies that

$$2 + A_n > q_n^{-2}2^{q_n}$$

for infinitely many n, and hence

$$q_n^2 A_n/2^{q_n} > 1/2$$

for infinitely many n, so $\theta \in \Theta_0$.

This means that $\Theta_0 \cup \mathbb{Q}$ contains V, and, therefore, is residual. Hence Θ_0 is residual (since $\mathbb{Q}$ is countable). This completes the proof.

Remarks. (1) The Proposition says that g is either not differentiable at "most" points or infinitely differentiable at "most" points, according to whether "most" is interpreted in the sense of category or measure. This is related to the well-known dichotomy between the Diophantine irrationals and the Liouville irrationals (those which are not Diophantine). See **[9]** for more on this interesting topic.

(2) Suppose we alter the definition of g so that 2^q is replaced by $w(q)$, where $w\colon \mathbb{Z}^+ \to \mathbb{Z}^+$ is some increasing function. Then the following are left to the reader. (See **[8]** for (a) and other related results.)

(a) If $w(q) = q^2$, then g is nowhere differentiable. (Use (2).)

(b) If $w(q) = q^3$, g is differentiable on a dense, uncountable set of irrationals, but nowhere twice differentiable.

(c) No matter how rapidly w increases, the set Θ_0 of points of nondifferentiability is residual.

As a consequence of (c), no function vanishing at the irrationals and discontinuous at the rationals can be differentiable at the irrationals. In fact, a little more argument shows that no function can be discontinuous at every rational but differentiable at every irrational. (This has been known, by another method of proof, for some time, e.g. **[3]**, **[2]**.) The following theorem implies (c) and the above statements, and provides a nice application of the Diophantine approximation point of view. (A slightly weaker version appears in **[6]** and is considered from a more general viewpoint in **[1]**.)

DEFINITION. *For* $\alpha > 0$, *a function* f *is said to be* α*-Hölder at* $x \in \mathbb{R}$ *if, for some* $M > 0$, $|f(x) - f(y)| < M|x - y|^{\alpha}$ *for all* y *near* x.

THEOREM. *If* $f: \mathbb{R} \to \mathbb{R}$ *is discontinuous on a dense set* S, *then there is a residual set* $A \subset \mathbb{R}$ *with the property that for all* $x \in A$ *and all* $\alpha > 0$, f *fails to be* α*-Hölder at* x.

Proof. S has a countable dense subset $\{r_n\}$. (Exercise.) The proof is inspired by the above proof that Θ_0 is residual, with $\{r_n\}$ playing the role of the rationals. Let $s: \mathbb{Z}^+ \to \mathbb{R} \cup \{\infty\}$ be defined by $s(n) = \limsup|f(r_n) - f(y)|$, where the lim sup is taken as $y \to r_n$. Then $s(n)$ is positive for all n since f is discontinuous on $\{r_n\}$. Let $h(n) = \min\{s(n), 1\}$. Fix $\alpha > 0$. For each n, $k \in \mathbb{Z}^+$, let $\delta(n, k) = k^{-1}h(n)^{1/\alpha}$. Let $A(k) = \bigcup_n B_{\delta(n, k)}(r_n)$. ($B_a(b)$ denotes the open ball of radius a, center b). This is open and dense in $\mathbb{R}$. Let $A_\alpha = \bigcap_k A(k)$. Pick $x \in A_\alpha$, and assume x is not in $\{r_n\}$ (otherwise f is not even continuous at x).
Then for every k there exists $r_{n(k)}$ such that $x \in B_{\delta(n(k),k)}(r_{n(k)})$, i.e., $|x - r_{n(k)}| < k^{-1}h(n(k))^{1/\alpha}$ (by definition of δ). Choose y_k appropriately near (possibly equal) to $r_{n(k)}$ so that

$$|y_k - r_{n(k)}| < |x - r_{n(k)}| \text{ and } |f(x) - f(y_k)| > \tfrac{1}{2}h(n(k)).$$

Then $|x - y_k| < 2|x - r_{n(k)}|$, so

$$\frac{|f(x) - f(y_k)|}{|x - y_k|^{\alpha}} > \frac{\frac{1}{2}h(n(k))}{\left(2|x - r_{n(k)}|\right)^{\alpha}} > \frac{k^{\alpha}h(n(k))}{2^{\alpha+1}h(n(k))} = k^{\alpha}/2^{\alpha+1},$$

and this tends to infinity as $k \to \infty$ (hence as $y_k \to x$). Hence f is not α-Hölder at x. Letting $A = \bigcap_j A_{1/j}$ completes the proof.

Final Remark. A version of the Proposition is also true in this more general context: for any countable dense set S there is a function $g: \mathbb{R} \to \mathbb{R}$ discontinuous on S and satisfying the conclusions of the Proposition (with respect to $\mathbb{R} \setminus S$ instead of $\mathbb{R} \setminus \mathbb{Q}$). Details will appear elsewhere.

(*Acknowledgement.* I thank the referees for helpful suggestions and for pointing out many of the cited articles.)

REFERENCES

1. E. M. Beesley, A. P. Morse, and D. C. Pfaff, Lipschitzian points, this MONTHLY, 79 (1972) 603–608.
2. R. P. Boas, A Primer of Real Functions, 3rd edition, Mathematical Association of America, Washington, D.C., 1981.
3. M. K. Fort, Jr., A theorem concerning functions discontinuous on a dense set, this MONTHLY, 58 (1951) 408–410.
4. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, 4th edition, Oxford University Press, London, 1960.
5. G. A. Heuer and the Undergraduate Research Participation Group, Functions continuous at the irrationals and discontinuous at the rationals, this MONTHLY, 72 (1965) 370–373.
6. G. A. Heuer, A property of functions discontinuous on a dense set, this MONTHLY, 73 (1966) 378–379.
7. W. J. LeVeque, Fundamentals of Number Theory, Addison-Wesley, Reading, 1977.
8. J. E. Nymann, An application of Diophantine approximation, this MONTHLY, 76 (1969) 668–671.
9. J. C. Oxtoby, Measure and Category, Springer-Verlag, New York, 1971.

EDITED BY JOAN P. HUTCHINSON AND STAN WAGON

More on the Fundamental Theorem of Calculus

CHARLES SWARTZ
Department of Mathematics, New Mexico State University, Las Cruces, NM 88003

and

BRIAN S. THOMSON
Department of Mathematics, Simon Fraser University, Burnaby, B. C., Canada V5A 1S6

In a note [1] in the MONTHLY, Botsko and Gosser point out that the standard version of the Fundamental Theorem of Calculus holds when the usual derivative is replaced by the right-hand derivative. We would like to point out that by making a *slight* alteration in the usual definition of the Riemann integral, we can obtain an integral for which the Fundamental Theorem of Calculus holds in *full generality*.

We begin by recalling one of the common definitions of the Riemann integral. If $P = \{a = x_0 < x_1 < \cdots < x_n = b\}$ is a partition of $[a, b]$, the *mesh* of P is $\max\{x_i - x_{i-1} : i = 1, \dots, n\}$.

DEFINITION 1. A function $f : [a, b] \to \mathbb{R}$ is *Riemann integrable* over $[a, b]$ if there exists $A \in \mathbb{R}$ such that for every $\varepsilon > 0$ there exists $\delta > 0$ such that if P is a partition of mesh less than δ and if $t_i \in [x_{i-1}, x_i]$, then

$$\left| \sum_{i=1}^{n} f(t_i)(x_i - x_{i-1}) - A \right| < \varepsilon.$$

The number A is called the *Riemann integral* of f and is denoted by $\int_a^b f$.

In order for a function $f : [a, b] \to \mathbb{R}$ to be Riemann integrable it is necessary that whenever the interval $[a, b]$ is partitioned into subintervals of length less than δ, the Riemann sums $\sum_{i=1}^{n} f(t_i)(x_i - x_{i-1})$ approximate the integral of f within ε. It is this requirement of being able to uniformly partition the interval that limits the scope of the Riemann integral. It would be much more desirable to somehow allow "variable length" partitions; for example, if one were attempting to approximate the area under the graph of $f(x) = 1/\sqrt{x}$, $0 < x \leq 1$, it would be natural to take the subintervals in an approximating partition to be very fine near the singularity $x = 0$. By making a slight modification in the definition above, we can easily achieve the ability to employ such variable length partitions.

First, note that the requirement in Definition 1 that the partition P have mesh less than δ can be replaced by the condition:

$$[x_{i-1}, x_i] \subseteq \left(t_i - \frac{\delta}{2}, t_i + \frac{\delta}{2} \right) \quad \text{where} \quad t_i \in [x_{i-1}, x_i]. \tag{1}$$

Now we can achieve the desired variable length partition by merely replacing the constant δ in (1) by a positive-valued function $\delta : [a, b] \to \mathbb{R}$, i.e., we replace (1) by:

$$[x_{i-1}, x_i] \subseteq (t_i - \delta(t_i), t_i + \delta(t_i)) \quad \text{where} \quad t_i \in [x_{i-1}, x_i]. \tag{1'}$$

By varying $\delta(t)$ as t varies along the interval $[a, b]$, the lengths of the subintervals in the partition satisfying (1′) will vary with $\delta(t_i)$.

We give the formal definition of the resulting integral. A tagged partition of $[a, b]$ is a finite set $T = \{x_0, x_1, \ldots, x_n; t_1, t_2, \ldots, t_n\}$ such that $\{x_0, x_1, \ldots, x_n\}$ is a partition of $[a, b]$ and $t_i \in [x_{i-1}, x_i]$; the point t_i is said to be a *tag* for the subinterval $[x_{i-1}, x_i]$. Any positive-valued function $\delta : [a, b] \to \mathbb{R}$ is called a *gauge* on $[a, b]$. We say that a tagged division T is *δ-fine* if (1′) is satisfied.

DEFINITION 2. A function $f : [a, b] \to \mathbb{R}$ is *gauge-integrable* over $[a, b]$ if there exists $A \in R$ such that for every $\varepsilon > 0$ there exists a gauge δ on $[a, b]$ such that if T is a δ-fine tagged partition of $[a, b]$, then

$$\left| \sum_{i=1}^{n} f(t_i)(x_i - x_{i-1}) - A \right| < \varepsilon.$$

The number A is called the (gauge) integral of f and is denoted by $\int_a^b f$. The gauge integral is also referred to as the generalized Riemann integral **[9]** or the Riemann complete integral **[5]**. From Definition 1, we see that a function is Riemann integrable iff it is gauge integrable with respect to constant-valued gauges. Note also that if a function is gauge integrable then its integral is the limit of a sequence of Riemann sums. One technicality must be taken care of in order for Definition 2 to make sense: it must be shown that every gauge δ has at least one δ-fine tagged partition. Currently this observation is ascribed to Pierre Cousin **[2]** but it may go somewhat earlier. The lemma has a curious habit of rediscovery: for example, each of the articles **[3]**, **[7]**, **[12]**, **[14]** contains a fresh account with similar applications. The proof requires only a compactness argument (based on the Bolzano-Weierstrass or Heine-Borel theorems) and indeed the lemma is equivalent to these theorems. The reader can find an elementary proof in **[9]**.

Before proceeding to the Fundamental Theorem of Calculus, consider the integrability of the Dirichlet function: $f(x) = 1$ for $0 \le x \le 1$ and rational and $f(x) = 0$ for $0 \le x \le 1$ and irrational. This is the most common example given of a bounded function that is not Riemann integrable, and will, therefore, furnish a comparison of the gauge and Riemann integrals. Let $\varepsilon > 0$ be given and for x irrational, set $\delta(x) = 1$. Let $\{z_i\}_{i=1}^{\infty}$ be an enumeration of the rationals in $[0,1]$ and set $\delta(z_i) = \varepsilon/2^{i+1}$. Now suppose that $T = \{x_0, x_1, \ldots, x_n; t_1, t_2, \ldots, t_n\}$ is a δ-fine tagged division of $[0,1]$. If t_i is not rational, the term $f(t_i)(x_i - x_{i-1})$ in the Riemann sum of f with respect to T is 0; if t_i is rational and $t_i = z_j$, then the term $f(t_i)(x_i - x_{i-1})$ in the Riemann sum is less than $2\delta(z_j) = \varepsilon/2^{j+1}$. Thus, we have

$$\left| \sum_{i=0}^{n-1} f(t_i)(x_i - x_{i-1}) \right| < 2 \sum_{j=1}^{\infty} \varepsilon/2^{j+1} = \varepsilon,$$

where the factor 2 is necessary since each z_j may be the tag for two of the subintervals in the partition. This implies that f is gauge integrable with $\int_0^1 f = 0$. Note that the gauge δ is definitely not a constant-valued gauge.

We now show that the Fundamental Theorem of Calculus is valid for the gauge integral in full generality. For this we require the following lemma which plays the role of the Mean Value Theorem in the usual proofs of the Fundamental Theorem of Calculus for the Riemann integral.

LEMMA 3 (STRADDLE LEMMA). *Let $F:[a, b] \to \mathbb{R}$ be differentiable at $z \in [a, b]$. Then for each $\varepsilon > 0$, there is a $\delta > 0$ such that*

$$|F(v) - F(u) - F'(z)(v - u)| \leq \varepsilon(v - u),$$

whenever $u \leq z \leq v$ and $[u, v] \subseteq [a, b] \cap (z - \delta, z + \delta)$.

Proof. Since F is differentiable at z, there is a $\delta > 0$ such that

$$|(F(x) - F(z))/(x - z) - F'(z)| < \varepsilon$$

for $0 < |x - z| < \delta$, $x \in [a, b]$. If $z = u$ or $z = v$, the conclusion is immediate so suppose $u < z < v$. Then,

$$\begin{aligned}
&|F(v) - F(u) - F'(z)(v - u)| \\
&\quad \leq |F(v) - F(z) - F'(z)(v - z)| + |F(z) - F(u) - F'(z)(z - u)| \\
&\quad < \varepsilon(v - z) + \varepsilon(z - u) = \varepsilon(v - u).
\end{aligned}$$

The geometric interpretation of the Straddle Lemma is clear. If the points u and v "straddle" z, then the slope of the chord between the points $(u, f(u))$ and $(v, f(v))$ is close to the slope of the tangent line at $(z, f(z))$.

THEOREM 4 (FUNDAMENTAL THEOREM OF CALCULUS). *If $F:[a, b] \to \mathbb{R}$ is differentiable on $[a, b]$, then F' is gauge integrable over $[a, b]$ and $\int_a^b F' = F(b) - F(a)$.*

Proof. Let $\varepsilon > 0$. For $z \in [a, b]$, let $\delta(z) > 0$ be the δ given by the Straddle Lemma. Suppose $T = \{x_0, x_1, \ldots, x_n; t_1, t_2, \ldots, t_n\}$ is a δ-fine tagged partition of $[a, b]$. Then by the Straddle Lemma,

$$\begin{aligned}
&\left| \sum_{i=1}^{n} F'(t_i)(x_i - x_{i-1}) - (F(b) - F(a)) \right| \\
&\qquad = \left| \sum_{i=1}^{n} \{F'(t_i)(x_i - x_{i-1}) - (F(x_i) - F(x_{i-1}))\} \right| \\
&\qquad < \sum_{i=1}^{n} \varepsilon(x_i - x_{i-1}) = \varepsilon(b - a),
\end{aligned}$$

and the conclusion follows.

Note that in general the gauge δ constructed above is not a constant-valued gauge and depends on the differentiability properties of F.

The versions of the Fundamental Theorem of Calculus for both the Riemann and Lebesgue integrals require the hypothesis that the derivative F' is integrable; it is part of the conclusion of Theorem 4 that the derivative F' is gauge integrable. For example, the derivative of the function $F(t) = t^2 \cos(\pi/t^2)$, $0 < t \leq 1$, $F(0) = 0$ is gauge integrable, but is not integrable for either the Riemann or Lebesgue integrals.

To further illustrate the utility of the gauge integral, we now proceed to generalize Theorem 4 by allowing the function F to be nondifferentiable at a countable number of points.

THEOREM 5. *Let $F:[a, b] \to \mathbb{R}$ be differentiable except perhaps at countably many points of $[a, b]$. Let $G:[a, b] \to \mathbb{R}$ be such that $G(x) = F'(x)$ when F is differentia-*

ble at x. If F is continuous, then G is gauge integrable over $[a, b]$ *with* $\int_a^b G = F(b) - F(a)$.

Proof. The proof is not substantially different and requires only some arithmetic to take care of the exceptional set $N = \{z_i : i = 1, 2, \dots\}$ where F may fail to be differentiable. Let $\varepsilon > 0$. For $x \notin N$ define $\delta(x) > 0$ by the Straddle Lemma as before. For $z_i \in N$, choose $\delta(z_i) > 0$ such that $|G(z_i)|2\delta(z_i) < \varepsilon/2^{i+2}$ and

$$|F(z_i) - F(z_i + h)| < \varepsilon/2^{i+3}$$

for $|h| \leq \delta(z_i)$ (by continuity). Suppose that $T = \{x_0, x_1, \dots, x_n; t_1, t_2, \dots, t_n\}$ is a δ-fine tagged partition of $[a, b]$. Then as before,

$$\left|\sum_{i=1}^{n} G(t_i)(x_i - x_{i-1}) - (F(b) - F(a))\right|$$
$$= \left|\sum_{i=1}^{n} \big(G(t_i)(x_i - x_{i-1}) - (F(x_i) - F(x_{i-1}))\big)\right|. \tag{2}$$

We break the sum on the right-hand side of (2) into two parts. Let Σ' denote the sum of the terms with tags $t_i \notin N$, and let Σ'' denote the sum of the terms with tags $t_i \in N$. As before the sum Σ' is less than $\varepsilon(b - a)$. For Σ'', if $t_i = z_j$, then

$$|G(t_i)(x_i - x_{i-1})| \leq |G(z_j)|2\delta(z_j) < \varepsilon/2^{j+2}$$

and

$$|F(x_i) - F(x_{i-1})| \leq |F(x_i) - F(z_j)| + |F(z_j) - F(x_{i-1})| < 2\varepsilon/2^{j+3}.$$

Hence,

$$\Sigma'' \leq 2\left(\sum_{j=1}^{\infty} \varepsilon/2^{j+2} + \sum_{j=1}^{\infty} \varepsilon/2^{j+2}\right) = \varepsilon,$$

where the factor of 2 accounts for the fact that each z_j may be the tag for at most two of the subintervals. Thus, the sum in (2) is less than $\varepsilon(b - a) + \varepsilon$ and the result follows.

Both versions of the Fundamental Theorem of Calculus given in Theorems 4 and 5 are well known for the gauge integral and can be found in [**9**]. There is also a divergence version of the Fundamental Theorem of Calculus for a gauge-type integral in n dimensions given in [**13**].

The functions $F(x) = 2x^{1/2}$, $0 \leq x \leq 1$, and $G(x) = x^{-1/2}$ for $0 < x \leq 1$ and $G(0) = 0$ provide a simple example where Theorem 5 is applicable but the more familiar version of the Fundamental Theorem of Calculus is not. Note that in this case, the integral

$$\int_0^1 x^{-1/2}\,dx = \int_0^1 G = \int_0^1 F' = 2$$

is computed without resorting to the limiting technique required by the Riemann approach.

Note that the continuity assumption in Theorem 5 is important. For example, if $F_1(x) = x + 1$ for $0 \leq x \leq 1$ and $F_1(x) = -x$ for $-1 \leq x < 0$, then $F_1'(x) = G(x)$ except for $x = 0$, but $\int_{-1}^1 G = 0$ while $F_1(1) - F_1(-1) = 1$.

As can be seen from the definition, the gauge integral has very much the same flavor as the Riemann integral, being obtained from a slight modification of the Riemann integral, and does not require a lot of technical apparatus for its introduction as is the case for the Lebesgue integral. However, despite the elementary appearance of the gauge integral, it leads to a very powerful theory of integration which encompasses the Riemann integral, the Cauchy-Riemann (improper Riemann) integral, and the Lebesgue integral. For this reason, the gauge integral would seem to be a very reasonable candidate for inclusion in an introductory real analysis course; it is as conceptually easy to describe as the Riemann integral and yet possesses all of the powerful properties of the Lebesgue integral including the Monotone and Dominated Convergence Theorems.

Remarkably, this simple modification of the Riemann integral was not introduced until approximately a century after Riemann's introduction of his integral in 1854. The gauge integral was independently introduced by Kurzweil **[6]** and Henstock **[4]**; Kurzweil used the integral to treat some questions in ordinary differential equations but did not develop any of the deep properties of the integral; Henstock established the convergence theorems for the integral.

The interested reader can find very readable expositions of the gauge integral in **[5]**, **[8]**, **[9]**. E. J. McShane also treats a "gauge-like" integral in **[10]**, **[11]**; he alters the definition above by dropping the requirement that the tag t_i belongs to its corresponding subinterval. The resulting integral is, surprisingly enough, exactly equivalent to the classical Lebesgue integral.

REFERENCES

1. M. W. Botsko and R. A. Gosser, Stronger versions of the Fundamental Theorem of Calculus, this MONTHLY, 93 (1986), 294–296.
2. P. Cousin, Sur les fonctions de n variables complexes, *Acta Math.*, 19 (1895) 1–62.
3. L. R. Ford, Interval additive propositions, this MONTHLY, 64 (1957) 106–108.
4. R. Henstock, Definitions of Riemann type of the variational integrals, *Proc. London Math. Soc.*, 11 (1961) 402–418.
5. ______, A Riemann-type integral of Lebesgue power, *Canadian J. Math*, 20 (1968) 79–87.
6. J. Kurzweil, Generalized ordinary differential equations and continuous dependence on a parameter, *Czech. Math. J.*, 82 (1957) 418–449.
7. J. Leinfelder, A unifying principle in real analysis, *Real Analysis Exchange*, 8 (1982/83) 511–518.
8. J. Mawhin, Introduction a l'Analyse, CABAY, Louvain-La-Neuve, 1983.
9. R. M. McLeod, The Generalized Riemann Integral, Carus Mathematical Monographs 20, MAA, 1980.
10. E. J. McShane, A unified theory of integration, this MONTHLY, 80 (1973) 349–359.
11. ______, Unified Integration, Academic Press, N.Y., 1983.
12. R. M. F. Moss and G. T. Roberts, A creeping lemma, this MONTHLY, 75 (1968) 649–652.
13. W. Pfeffer, The divergence theorem, *Trans. Amer. Math. Soc.*, 295 (1986) 665–687.
14. P. Shanahan, A unified proof of several basic theorems of real analysis, this MONTHLY, 79 (1972) 895–898.

A Strong Inverse Function Theorem

WILLIAM J. KNIGHT
Department of Mathematics and Computer Science, Indiana University, South Bend, IN 46634

A linear transformation from $\mathbb{R}^n$ into the same space $\mathbb{R}^n$ is one-to-one if and only if its matrix relative to, say, the standard basis has nonzero determinant. In this

case the transformation maps $\mathbb{R}^n$ *onto* $\mathbb{R}^n$. This much everyone knows. A trivial extension says that an affine function (a constant plus a linear transformation) from $\mathbb{R}^n$ into $\mathbb{R}^n$ is one-to-one if and only if the matrix of its linear part has nonzero determinant, in which case the function maps $\mathbb{R}^n$ onto $\mathbb{R}^n$.

Now consider a non-linear function $\vec{f} = (f_1, \ldots, f_n)$ from a subset of $\mathbb{R}^n$ into $\mathbb{R}^n$. We say $\vec{f}$ is *differentiable* at a point $\vec{a}$ iff $\vec{f}$ is "affine at $\vec{a}$," by which we mean there exist a linear transformation $\vec{T}$ and a "vanishing" function $\vec{E}$ such that

$$\vec{f}(\vec{x}) = \vec{f}(\vec{a}) + \vec{T}(\vec{x} - \vec{a}) + \|\vec{x} - \vec{a}\|\vec{E}(\vec{x})$$

for all $\vec{x}$ in some neighborhood of $\vec{a}$, and $\vec{E}(\vec{x}) \to \vec{0}$ as $\vec{x} \to \vec{a}$. If we write $\vec{c} = \vec{f}(\vec{a}) - \vec{T}(\vec{a})$, then $\vec{f}(\vec{x}) \approx \vec{c} + \vec{T}(\vec{x})$ (an affine function) when $\vec{x} \approx \vec{a}$. We call $\vec{T}$ the *total derivative* of $\vec{f}$ at $\vec{a}$. The determinant of its matrix is the Jacobian of $\vec{f}$ at $\vec{a}$. The property of affine maps mentioned in the first paragraph suggests that if this Jacobian is nonzero, then $\vec{f}$ will be one-to-one on some neighborhoods of $\vec{a}$ and map that neighborhood *onto* a neighborhood of $\vec{f}(\vec{a})$. This pleasant conjecture is defeated by the following function from $\mathbb{R}^2$ to $\mathbb{R}^2$:

$$f_1(x, y) = x + 2x^2 \sin(1/x) \quad \text{if} \quad x \neq 0, \qquad f_1(0, y) = 0,$$
$$f_2(x, y) = y.$$

This function is differentiable at $(0,0)$ and has Jacobian 1, but there is no neighborhood of $(0,0)$ on which it is one-to-one (for the proof of this assertion it helps to note that $(d/dx)\,(x + 2x^2 \sin(1/x))$ takes positive and negative values on every neighborhood of 0). This explains why texts that prove the Inverse Function Theorem must use stronger hypotheses. Typically they assume that $\vec{f}$ is differentiable throughout some neighborhood of $\vec{a}$, or even that $\vec{f}$ has continuous partial derivatives on such a neighborhood. The purpose of this paper is to prove that the classical sufficient condition for differentiability of $\vec{f}$ at a single point $\vec{a}$ is also sufficient to make $\vec{f}$ invertible on a neighborhood of $\vec{a}$ when the Jacobian is nonzero. The classical condition to which we refer is the following: *the first-order partial derivatives of the component functions f_i exist on a neighborhood of $\vec{a}$ and are continuous at $\vec{a}$.* In this case we shall say that "$\vec{f}$ is minimally C^1 at $\vec{a}$." It is our pious hope that textbook writers of the future, seeing the simplicity and strength of this condition, will say simply that "$\vec{f}$ is C^1 at $\vec{a}$."

ONE-POINT INVERSE FUNCTION THEOREM. *Let $\vec{f}$ be a function from a subset of $\mathbb{R}^n$ into the same space $\mathbb{R}^n$. Suppose $\vec{f}$ is minimally C^1 at a point $\vec{a}$ and the Jacobian of $\vec{f}$ at $\vec{a}$ is nonzero. Then there exists a neighborhood $\mathcal{N}$ of $\vec{a}$ such that $\vec{f}$ is one-to-one and continuous on $\mathcal{N}$ and $\vec{f}[\mathcal{N}]$ is a neighborhood of $\vec{f}(\vec{a})$. Let $\vec{g}$ denote the inverse function of the restriction of $\vec{f}$ to $\mathcal{N}$. Then $\vec{g}$ is continuous on $\vec{f}[\mathcal{N}]$ and differentiable at $\vec{f}(\vec{a})$, and the Jacobian matrix of $\vec{g}$ at $\vec{f}(\vec{a})$ is the inverse of the Jacobian matrix of $\vec{f}$ at $\vec{a}$.*

To prove this theorem it suffices to prove that the hypotheses imply that $\vec{f}$ is *strongly differentiable* at $\vec{a}$, which means (see [2], [3]) there is a linear transformation $\vec{T}: \mathbb{R}^n \to \mathbb{R}^n$ and a vanishing function $\vec{F}: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ such that

$$\vec{f}(\vec{x}) - \vec{f}(\vec{y}) = \vec{T}(\vec{x} - \vec{y}) + \|\vec{x} - \vec{y}\|\vec{F}(\vec{x}, \vec{y}) \tag{1}$$

for all $\vec{x}$ and $\vec{y}$ in some neighborhood of $\vec{a}$, and $\vec{F}(\vec{x}, \vec{y}) \to \vec{0}$ as $(\vec{x}, \vec{y}) \to (\vec{a}, \vec{a})$. As shown (tacitly) in [4], strong differentiability of $\vec{f}$ at $\vec{a}$ is enough to deliver most

of the conclusions of the One-Point Inverse Function Theorem stated above (in fact the inverse function $\vec{g}$ can even be shown to be *strongly* differentiable at $\vec{f}(\vec{a})$). Thus the proof of our theorem comes to rest on the following lemma.

LEMMA. *Let f be a function from a subset of $\mathbb{R}^n$ into $\mathbb{R}^n$ and let $\vec{a}$ be a point at which $\vec{f}$ is minimally C^1. Then $\vec{f}$ is strongly differentiable at $\vec{a}$. (Note: the proof does not require the domain and range spaces to have the same dimension.)*

Proof. While it is possible to prove that these conditions imply the hypotheses of a simple extension of Theorem 1 of [3, p. 972], it seems preferable to give a proof from first principles. Let $\vec{T}$ denote the total derivative of $\vec{f}$ at $\vec{a}$. This exists because $\vec{f}$ is minimally C^1 at $\vec{a}$. Let B be a box about $\vec{a}$ on which all first-order partial derivatives of the component functions f_i exist. Fix $\vec{x}$ and $\vec{y}$ in B. By moving from $\vec{x}$ to $\vec{y}$ along lines parallel to coordinate axes and applying the Mean Value Theorem for real-valued functions, one variable at a time, we obtain

$$f_i(\vec{x}) - f_i(\vec{y}) = \sum_{k=1}^{n} D_k f_i(\vec{c}_{ik})(x_k - y_k) \tag{2}$$

for each component function f_i and for appropriately chosen points $\vec{c}_{ik}$ along the path from $\vec{x}$ to $\vec{y}$. These points $\vec{c}_{ik}$ are forced toward $\vec{a}$ as $(\vec{x}, \vec{y}) \to (\vec{a}, \vec{a})$, so the difference defined by

$$E_{ik}(\vec{x}, \vec{y}) = D_k f_i(\vec{c}_{ik}) - D_k f_i(\vec{a})$$

approaches 0 as $(\vec{x}, \vec{y}) \to (\vec{a}, \vec{a})$. Thus we can write (2) in the form

$$\begin{aligned} f_i(\vec{x}) - f_i(\vec{y}) &= \sum_{k=1}^{n} D_k f_i(\vec{a})(x_k - y_k) + \sum_{k=1}^{n} E_{ik}(\vec{x}, \vec{y})(x_k - y_k) \\ &= \vec{\nabla} f_i(\vec{a}) \cdot (\vec{x} - \vec{y}) + \vec{E}_i(\vec{x}, \vec{y}) \cdot (\vec{x} - \vec{y}), \end{aligned} \tag{3}$$

where $\vec{E}_i = (E_{i1}, \dots, E_{in})$. The first term on the right side of (3) is the ith component of $\vec{T}(\vec{x} - \vec{y})$. To handle the second term we write

$$\vec{F}(\vec{x}, \vec{y}) = \frac{1}{\|\vec{x} - \vec{y}\|}\left(\vec{E}_1(\vec{x}, \vec{y}) \cdot (\vec{x} - \vec{y}), \dots, \vec{E}_n(\vec{x}, \vec{y}) \cdot (\vec{x} - \vec{y})\right)$$

if $\vec{x} \neq \vec{y}$ and set $\vec{F}(\vec{x}, \vec{x}) = \vec{0}$. This yields equation (1). Moreover,

$$\|\vec{F}(\vec{x}, \vec{y})\| \leq \frac{1}{\|\vec{x} - \vec{y}\|} \sum_{i=1}^{n} |\vec{E}_i(\vec{x}, \vec{y}) \cdot (\vec{x} - \vec{y})| \leq \sum_{i=1}^{n} \|\vec{E}_i(\vec{x}, \vec{y})\|$$

by the inequality $\|\vec{h}\| \leq \Sigma |h_i|$ and the Cauchy-Schwarz Inequality. It follows that $\vec{F}(\vec{x}, \vec{y}) \to \vec{0}$ as $(\vec{x}, \vec{y}) \to (\vec{a}, \vec{a})$. □

Some texts (e.g., [1]) prove that the following is a sufficient condition for (ordinary) differentiability of a real-valued function f of several variables: *all but one of the first-order partials of f exist on a neighborhood of $\vec{a}$ and are continuous at $\vec{a}$; the exceptional partial derivative exists at $\vec{a}$.* This result can easily be extended to vector functions. Does this condition also suffice for *strong* differentiability at $\vec{a}$? No. The function given as an example earlier satisfies this condition at (0, 0), but it cannot be strongly differentiable there because it has a nonzero Jacobian at (0, 0) and yet is not locally invertible there.

Acknowledgements. The author thanks Professors Alan Howard and Charles Stanton, and the referee, for several helpful suggestions.

REFERENCES

1. Tom M. Apostol, Mathematical Analysis, 2nd. ed., Addison-Wesley, Reading, Mass., 1974.
2. E. B. Leach, A note on inverse function theorems, *Proc. Amer. Math. Soc.*, 12 (1961) 694–697.
3. A. Nijenhuis, Strong derivatives and inverse mappings, *Amer. Math. Monthly*, 81 (1974) 969–980.
4. K. T. Smith, Primer of Mathematical Analysis, Springer-Verlag, New York, 1983.

A Constructive Proof of the Partial Fraction Decomposition

DAN SCOTT
Department of Mathematics, Rose-Hulman Institute of Technology, Terre Haute, IN 47803

DONALD R. PEEPLES
Department of Mathematics, Mary Washington College, Fredericksburg, VA 22401

The method of partial fraction decomposition is central to the utilization of rational functions in analysis (integration, differential equations, etc.). An efficient method of partial fraction decomposition appears in classical algebra texts such as Chrystal **[1]**; however, the efficiency of this "classical method" was demonstrated by Straight and Dowds only in 1984 **[2]**. The classical method was shown to work on all rational functions whose denominator contains at most one irreducible quadratic function. We show that the classical method can be modified to work on all rational functions while retaining the same efficiency; furthermore, our proof is *constructive*, is performed in a general setting, and uses techniques of modern algebra.

Straight and Dowds **[2]** point out that the classical method is recursive and that, whereas the usual method described in modern calculus texts requires $O(n^3)$ operations for a rational function whose denominator has degree n, the classical method takes only $O(n^2)$ operations and is "algebraically simpler"; furthermore, Chrystal points out that it is somewhat self-checking. All four of these properties are retained when the classical method is extended to include all rational functions.

Notation. The rational function that we wish to decompose will be denoted by p/q. Let $R[x]$ be the ring of polynomials over the reals; $R[x]/(f(x))$ is the quotient field, where f stands for some irreducible polynomial of $R[x]$. (Thus f is a linear or a quadratic polynomial.) It is well known that the elements of $R[x]/(f(x))$ can be represented by real polynomials of degree less than the degree of f. Also well known is that there exists a homomorphism φ of $R[x]$ into $R[x]/(f(x))$ defined by $\varphi(u) = u \bmod f$, where u is any element of $R[x]$. For the sake of clarity all arithmetic operations in the quotient field will be enclosed in a circle.

THEOREM (Partial fraction decomposition). *Suppose that the denominator of the rational function p/q has been factored so that $p/q = p/(wf^k)$ where f is an irreducible element of $R[x]$ and w and f are coprime. Then there exist unique polynomials s and t of $R[x]$ with degree$(s) <$ degree(f) such that*

$$\frac{p}{q} = \frac{p}{wf^k} = \frac{s}{f^k} + \frac{t}{wf^{k-1}}. \tag{1}$$

And in fact s and t are given by

$$s = \varphi(p) \oslash \varphi(w); \qquad t = (p - sw)/f. \tag{2a, b}$$

Comments: Note that s is actually thought of as both an element of $R[x]/(f(x))$ *and* as an element of $R[x]$ as appropriate. This is allowable in light of the earlier remark made about the usual representation of elements in $R[x]/(f(x))$. It should also be clear that if this theorem can be established, then it may be applied repeatedly to yield a complete partial fraction decomposition of p/q. (Further factorization of the denominator will most likely be necessary.)

Proof. Equation (1) is equivalent to

$$p = sw + ft. \tag{3}$$

Applying the homomorphism, φ, to (3) we obtain

$$\varphi(p) = \varphi(s) \otimes \varphi(w). \tag{4}$$

Since degree(s) < degree(f), $\varphi(s) = s$ so solving (4) for s we obtain

$$s = \varphi(p) \oslash \varphi(w)$$

as desired. Thus *if* (1) holds then (2a) must follow. Then (2b) follows from (3). Thus we have now established the *uniqueness* of the decomposition. The only additional thing that must be verified is that t, as defined by (2b), is actually a *polynomial*, that is, that $p - sw$ is a multiple of f. To do this simply observe that

$$\varphi(p - sw) = \varphi(p) \ominus [\varphi(p) \oslash \varphi(w)] \otimes \varphi(w) = 0. \tag{5}$$

This completes the proof. It should be noted that one may find s and t without ever dealing explicitly with operations in $R[x]/(f(x))$, simply by solving the equation $p \bmod f = sw \bmod f$ obtained from (3). In practice this seems to be the easiest way to proceed.

Analysis of Complexity. In order to facilitate easy comparison of the complexity of our algorithm with other methods, we shall assume that the degree of the numerator of the rational function is less than the degree of the denominator. The algorithm can, however, be applied even if this is not the case.

Suppose the denominator of p/q has been completely factored into the product of powers of linear and irreducible quadratic terms. Let m be the sum of the powers of these terms, and n the degree of the denominator. Then it must be that $n/2 \leq m \leq n$. The number of times that the theorem must be applied, in order to completely decompose p/q, is $m - 1$. Thus the number of applications of the theorem is $O(n)$. Each time the theorem is applied it is necessary to perform the following steps: 1) compute $\varphi(p)$ and $\varphi(w)$—each of which may be done by division and requires no more than $O(n)$ steps; 2) compute $\varphi(p) \oslash \varphi(w)$—which takes a small fixed number of operations; and 3) compute $t = (p - sw)/f$ by division—which requires no more than $O(n)$ steps. Thus the theorem is applied $O(n)$ times and each time requires $O(n)$ operations; therefore the decomposition requires $O(n^2)$ operations overall. More precise estimates may be obtained by analyzing the algorithm in more detail.

Example. A short example should help clarify the discussion. We shall find the partial fraction decomposition of $(x^3 - x + 4)/(x^2 + 1)(x - 1)^2$. The first applica-

tion of the theorem gives us

$$\frac{x^3 - x + 4}{(x^2+1)(x-1)^2} = \frac{s}{x^2+1} + \frac{t}{(x-1)^2}, \tag{6}$$

where

$$s = \varphi(p) \oslash \varphi(w) = (-2x+4) \oslash (-2x) = 2x+1 \tag{7}$$

and

$$t = (p - sw)/f = \left[(x^3 - x + 4) - (2x+1)(x-1)^2\right]/(x^2+1) = -x+3. \tag{8}$$

Next, we use the theorem to decompose $(-x+3)/(x-1)^2$. We obtain

$$\frac{-x+3}{(x-1)^2} = \frac{s}{(x-1)^2} + \frac{t}{x-1}, \tag{9}$$

where

$$s = \varphi(p) \oslash \varphi(w) = 2/1 = 2 \tag{10}$$

and

$$t = (p - sw)/f = [(-x+3) - (2)(1)]/(x-1) = -1. \tag{11}$$

So, putting everything together, we have

$$\frac{x^3 - x + 4}{(x^2+1)(x-1)^2} = \frac{2x+1}{x^2+1} + \frac{2}{(x-1)^2} - \frac{1}{x-1}. \tag{12}$$

The approach used above is certainly not the most efficient implementation for hand calculations, though it does have the advantage of staying close to the proof of the theorem. A more elementary version of this method *can* be taught to freshmen calculus students (see the note after the proof). In fact, in our experience, they seem to prefer it to the usual method of equating coefficients. This is partly because of the smaller number of calculations, and partly because after each fraction in the expansion is found, there is a division that helps determine if that term was found correctly. The divisibility check is made when it is found that *f divides evenly* into $p - sw$.

REFERENCES

1. G. Chrystal, Textbook of Algebra (part 1), Dover Publications, New York, 1961, pp. 151–159.
2. H. Joseph Straight and Richard Dowds, An alternate method for finding the partial fraction decomposition of a rational function, *Amer. Math. Monthly*, 91(1984) 365–367.
3. B. L. van der Waerden, Modern Algebra (trans. by F. Blum), Frederich Ungar Publishing, New York, 1949, pp. 88–90.

Let $a \bmod m$ denote the arithmetic progression $a + km$: $k \in \mathrm{Z}$. Suppose $b \bmod n$ is disjoint from $a \bmod m$. What is the smallest positive integer r such that there is an arithmetic progression $c \bmod r$ which is disjoint from both $a \bmod m$ and $b \bmod n$?

E 3279. *Proposed by Stanley Rabinowitz, Alliant Computer Systems Corp., Littleton, MA.*

For some n find a simplex in E^n with integer edges and volume 1. (Cf. Problem 1261 in *Mathematics Magazine*, [1987, p. 40; 1988, p. 118].)

E 3280. *Proposed by Peter J. Ferraro, Roselle Park, NJ.*

Let (X, T) be a topological space such that $\{p\}'$ is closed for every point $p \in X$. Prove that A' is closed for every subset A of X. Here A' is the set of points $x \in X$ such that every open set containing x contains a point of $A \setminus \{x\}$.

E 3281. *Proposed by Sandi Klavžar and Marko Petkovšek, University of Ljubljana, Yugoslavia.*

Given a graph G, let $C(G)$ be the inclusion ordering on the vertex subsets of G inducing connected subgraphs. Prove that $C(G)$ is a lattice if and only if G is connected and every cycle of G induces a complete subgraph of G.

SOLUTION OF ELEMENTARY PROBLEMS

More on Arithmetic-Geometric-Harmonic Means

E 3141 [1986, 298]. *Proposed by Mo Song-Qing, The Beijing 19th Middle School, Beijing, China.*

Suppose $a_i > 0$, $i = 1, \dots, n$. Consider the inequality

$$\frac{a_1 + a_2 + \cdots + a_k}{k} \cdot \frac{a_2 + a_3 + \cdots + a_{k+1}}{k} \cdots \frac{a_n + a_1 + \cdots + a_{k-1}}{k}$$
$$\geqslant \frac{a_1 + a_2 + \cdots + a_n}{n} \cdot (a_1 a_2 \cdots a_n)^{(n-1)/n} \tag{1}$$

for $2 \leqslant k \leqslant n - 1$.

(a) Prove or disprove (1) for $n = 3$.

(b)* Prove or disprove (1) for $n > 3$.

Solution I to (a) by Michael Vowe, Therwil, Switzerland. Let

$$f(x, y, z) = \frac{1}{8}(x + y)(y + z)(z + x) - \frac{1}{3}(x + y + z)(xyz)^{2/3}.$$

We show $f(x, y, z) \geqslant 0$ for positive x, y, z, with equality if and only if $x = y = z$. Since $f(tx, ty, tz) = t^3 f(x, y, z)$, we may assume $x + y + z = 1$. In this case, we want to prove

$$(1 - x)(1 - y)(1 - z) \geqslant \frac{8}{3}(xyz)^{2/3},$$

which holds if and only if

$$(xyz)^{1/3} \geqslant \frac{8}{3} \cdot \frac{1}{1/x + 1/y + 1/z - 1} \tag{2}$$

Because $x + y + z = 1$, the arithmetic-harmonic mean inequality becomes $1/3 \geqslant 3/(1/x + 1/y + 1/z)$. This rearranges to $9(1/x + 1/y + 1/z - 1) \geqslant 8(1/x + 1/y + 1/z)$, or

$$\frac{3}{1/x + 1/y + 1/z} \geqslant \frac{8}{3} \cdot \frac{1}{1/x + 1/y + 1/z - 1} \tag{3}$$

The inequality between the geometric and harmonic means implies that the left side of (2) is as large as the left side of (3), which proves (2) and thus (1).

Solution II to (a) by László Cseh, Odorheiu-Secuiesc, Romania. For $n = 3$, we prove a sharpened form of (1), namely,

$$\frac{a_1 + a_2}{2} \cdot \frac{a_2 + a_3}{2} \cdot \frac{a_3 + a_1}{2} \geqslant \left(\frac{a_1 + a_2 + a_3}{3}\right)^{3/2} (a_1 a_2 a_3)^{1/2}. \tag{4}$$

For $n = 3$, (1) follows from (4) by the arithmetic-geometric mean inequality.

(4) follows from an inequality of T. Popoviciu (in "Sur certaines inégalités qui caractérisent les fonctions convexes," Analele Stiintifice Univ. "Al. I. Cuza" Iasi Sect. I-a Mat. (N.S.) 11B(1965) 155–164). If f is a convex real-valued function on $[a, b]$ and $x, y, z \in [a, b]$, Popoviciu proved that

$$f\left(\frac{x + y + z}{3}\right) + \frac{f(x) + f(y) + f(z)}{3} \geqslant \frac{2}{3}\left[f\left(\frac{x + y}{2}\right) + f\left(\frac{y + z}{2}\right) + f\left(\frac{z + x}{2}\right)\right]. \tag{5}$$

The function $f(x) = -\log x$ is convex on $(0, +\infty)$. Putting $x = a_1$, $y = a_2$, $z = a_3$ and exponentiating Popoviciu's inequality yields (4).

Editorial comment. Murray Klamkin remarked that, by an argument similar to that in Cseh's solution, Popoviciu's more general inequality for convex functions

$$(n - 2)f\left(\frac{s}{n}\right) + \frac{1}{n} \sum_{j=1}^{n} f(x_j) \geqslant \frac{n - 1}{n} \sum_{j=1}^{n} f\left(\frac{s - x_j}{n - 1}\right),$$

where $s = \sum_{j=1}^{n} x_j$, implies the truth of (1) for the case $k = n - 1$, for any $n \geqslant 3$.

The truth of (1) in the case $k = n - 1$ was also established by M. J. Pelling. In addition Pelling showed that (1) holds in the case $k = 2$, $n = 4$, so that (1) is valid for $n = 4$ as well as for $n = 3$.

On the other hand Pelling showed that, for each fixed $k \geqslant 2$, the assertion (1) is false when

$$n > \frac{\pi^2}{3}(k^2 - 1).$$

In particular, if $k = 2$, (1) is false when $n \geqslant 10$.

Pelling's elegant argument for the falsity of (1) when $n > \pi^2(k^2 - 1)/3$ is as follows.

Suppose k is fixed, $k \geqslant 2$, and $n > k$. Under the convention that $a_{n+i} = a_i$, (1) may be written as

$$L \leqslant R, \tag{6}$$

where

$$L = -\sum_{j=1}^{n} \log\left(\frac{a_j + a_{j+1} + \cdots + a_{j+k-1}}{k}\right),$$

$$R = -\log\left(\frac{1}{n}\sum_{j=1}^{n} a_j\right) - \frac{n-1}{n}\sum_{j=1}^{n} \log a_j.$$

If (6) were universally valid, it would hold in particular when $a_j = 1 + \varepsilon\cos(2\pi j/n)$ for $1 \leqslant j \leqslant n$. For this choice of the a_j we have

$$\begin{aligned} L &= \sum_{j=1}^{n} \frac{\varepsilon^2}{2k^2}\left(\sum_{r=0}^{k-1} \cos\frac{2\pi(j+r)}{n}\right)^2 + O(\varepsilon^3) \\ &= \frac{\varepsilon^2 n}{4k} + \frac{\varepsilon^2 n}{2k^2}\sum_{0 \leqslant r < s < k} \cos\frac{2\pi(s-r)}{n} + O(\varepsilon^3), \end{aligned}$$

where we have used the identity

$$\sum_{j=1}^{n} \cos\frac{2\pi(j+r)}{n}\cos\frac{2\pi(j+s)}{n} = \frac{n}{2}\cos\frac{2\pi(s-r)}{n} \qquad (n \geqslant 3).$$

Similarly

$$\begin{aligned} R &= \frac{n-1}{n}\sum_{j=1}^{n} \frac{\varepsilon^2}{2}\cos^2\frac{2\pi j}{n} + O(\varepsilon^3) \\ &= \frac{(n-1)\varepsilon^2}{4} + O(\varepsilon^3). \end{aligned}$$

Thus a necessary condition for (6) is that

$$\frac{n}{4k} + \frac{n}{2k^2}\sum_{0 \leqslant r < s < k} \cos\frac{2\pi(s-r)}{n} \leqslant \frac{(n-1)}{4}$$

or

$$\frac{k^2}{2n} \leqslant \frac{1}{2}k(k-1) - \sum_{0 \leqslant r < s < k} \cos\frac{2\pi(s-r)}{n}$$

or

$$\frac{k^2}{2n} \leqslant \sum_{0 \leqslant r < s < k}\left(1 - \cos\frac{2\pi(s-r)}{n}\right) = \sum_{0 \leqslant r < s < k} 2\sin^2\frac{\pi(s-r)}{n}.$$

Using the inequality $|\sin x| \leqslant |x|$, we obtain

$$\frac{k^2}{2n} \leqslant \sum_{0 \leqslant r < s < k} \frac{2\pi^2(s-r)^2}{n^2} = \sum_{j=1}^{k-1} (k-j)\frac{2\pi^2 j^2}{n^2}$$

$$= \frac{2\pi^2}{n^2}\left\{\frac{k^2(k-1)(2k-1)}{6} - \frac{k^2(k-1)^2}{4}\right\} = \frac{\pi^2}{n^2}\frac{k^2(k^2-1)}{6}.$$

Thus a necessary condition for the validity of (1) or (6) is that

$$n \leqslant \frac{\pi^2}{3}(k^2-1).$$

Part (a) was solved also by O. Krafft (West Germany), J. E. Pečarić, and the proposer.

Weitzenböck Generalized

E 3150 [1986, 400]. *Proposed by George A. Tsintsifas, Thessaloniki, Greece.*

Let ABC be a triangle with sides a, b, c and area F. It is well known that $a^2 + b^2 + c^2 \geqslant 4F\sqrt{3}$. If p, q, r are arbitrary positive real numbers, prove that

$$\frac{p}{q+r}a^2 + \frac{q}{r+p}b^2 + \frac{r}{p+q}c^2 \geqslant 2F\sqrt{3}.$$

Solution I by Victor Pambuccian, Bucharest, Romania. Put $a = a_1$, $b = a_2$, $c = a_3$, $p = p_1$, $q = p_2$, $r = p_3$, and $k = \Sigma p_i$. The left side of the proposed inequality becomes

$$\Sigma\frac{p_i a_i^2}{k-p_i} = k\Sigma\frac{a_i^2}{k-p_i} - \Sigma a_i^2 = \tfrac{1}{2}\Sigma(k-p_i)\Sigma\frac{a_i^2}{k-p_i} - \Sigma a_i^2 \geqslant \tfrac{1}{2}(\Sigma a_i)^2 - \Sigma a_i^2,$$

where the inequality in the last step follows from Cauchy's Inequality.

Furthermore,

$$\frac{1}{2}(\Sigma a_i)^2 - \Sigma a_i^2 = \frac{1}{2}(\Sigma 2l/a_i - \Sigma a_i^2),$$

where $l = a_1 a_2 a_3$. Let A_i denote the angle opposite a_i. Since $\Sigma a_i^2 = \Sigma(2l/a_i)\cos A_i$, we can rewrite the desired expression as

$$\frac{1}{2}(\Sigma a_i)^2 - \Sigma a_i^2 = \frac{1}{2}\Sigma\frac{2l}{a_i}(1-\cos A_i) = \Sigma\frac{l}{a_i}2\sin^2\frac{A_i}{2}$$

$$= \Sigma\frac{2F}{\sin A_i}2\sin^2\frac{A_i}{2} = 2F\Sigma\tan\frac{A_i}{2}.$$

We now need only show $\Sigma\tan(A_i/2) \geqslant \sqrt{3}$. This follows from the convexity of the tangent function on the interval $(0, \pi/2)$.

Solution II by O. P. Lossers, Eindhoven University of Technology, The Netherlands. With the notation p_i, a_i as above, we may assume that $\Sigma p_i = 1$. Using Lagrange multipliers for the function $f(p_1, p_2, p_3) = \Sigma p_i a_i^2/(1-p_i)$ on the set

$\{(p_1, p_2, p_3)\colon p_i > 0, \Sigma p_i = 1\}$, we find a minimum with $p_i = (s - a_i)/s$, where $s = (1/2)\Sigma a_i$. Evaluating f at this point, we must prove that $\Sigma(s - a_i)a_i \geqslant 2F\sqrt{3}$.

Because both sides of this inequality are positive, we may check the inequality by taking the difference of their squares. To express both sides in convenient forms, we express the side-lengths as $a_1 = y + z$, $a_2 = z + x$, $a_3 = x + y$. Then $\Sigma(s - a_i)a_i = 2(xy + yz + zx)$ and $F^2 = s(s - a_1)(s - a_2)(s - a_3) = xyz(x + y + z)$. Thus

$$\begin{aligned}[\Sigma(s - a_i)a_i]^2 - 12F^2 &= 4\left[(xy + yz + zx)^2 - 3xyz(x + y + z)\right]\\ &= 2x^2(y - z)^2 + 2y^2(z - x)^2 + 2z^2(x - y)^2 \geqslant 0.\end{aligned}$$

Furthermore, equality occurs if and only if $x = y = z$, i.e., $a_1 = a_2 = a_3$.

Solution III and generalization by M. S. Klamkin, University of Alberta, Canada. Again with the same notation, we derive more generally an inequality for $S = \Sigma p_i a_i^{4n}/(k - p_i)$, where $1 \geqslant n \geqslant 0$ and $a_i \geqslant 0$. By Cauchy's Inequality, we have $2(S + \Sigma a_i^{4n}) = \Sigma(k - p_i)\Sigma a_i^{4n}/(k - p_i) \geqslant (\Sigma a_i^{2n})^2$. Letting $R = \Sigma a_i^n$, this is equivalent to

$$2S \geqslant R\Pi(R - 2a_i^n) \tag{1}$$

with equality if and only if $a_1^{2n}/(p_2 + p_3) = a_2^{2n}/(p_3 + p_1) = a_3^{2n}/(p_1 + p_2)$. If the $\{a_i^n\}$ are not the sides of a triangle, then the right side of (1) is negative. Suppose that $\{a_i\}$ form a triangle and that $1 \geqslant n \geqslant 0$. Then $\{a_i^n\}$ also form a triangle, whose area we denote by F_n. The right side of (1) is 16 times the square of the side-length formula for area, yielding $2S \geqslant 16F_n^2$.

We now use Oppenheim's generalization [**1**] of the Finsler-Hadwiger inequality, i.e., $4F_n/\sqrt{3} \geqslant (4F/\sqrt{3})^n$ for $1 \geqslant n \geqslant 0$, with equality for $n < 1$ if and only if $a_1 = a_2 = a_3$. With (1), this yields the generalization

$$2S \geqslant 3\left(4F/\sqrt{3}\right)^{2n} \qquad (0 \leqslant n \leqslant 1). \tag{2}$$

The proposed inequality corresponds to the special case $n = 1/2$. The special case $n = 1$ corresponds to the proposer's problem #1051 in *Crux Mathematicorum*, 11(1985) 187.

1. A. Oppenheim, Inequalities involving elements of triangles, quadrilaterals, or tetrahedra, *Publications de la Faculté d'Electrotechnique de l'Université à Belgrade*, No. 461–497(1974) 257–263.

Also solved by P. Dwinger, J. Fukuta (Japan), J. V. Harrell, W. Janous (Austria), D. C. Kay, K. Schilling, South Alabama Problem Group, M. Vowe (Switzerland), R. C. Vrem, and the proposer. Partially solved by H. Geiges (West Germany), L. Kuipers (Switzerland), and I. A. Sakmar (Turkey). The inequality $a^2 + b^2 + c^2 \geqslant 4F\sqrt{3}$ goes back to Weitzenböck, *Math. Z.*, 5(1919) 137–146.

An Inscribed Triangle Inequality

E 3154 [1986, 482]. *Proposed by George A. Tsintsifas, Thessaloniki, Greece.*

Let A_1, B_1, C_1 be points on the sides a, b, c of the triangle ABC, respectively, and let a_1, b_1, c_1 be the sides of the triangle $A_1B_1C_1$. Prove that $a^2b_1c_1 + b^2c_1a_1 + c^2a_1b_1 \geqslant 4F^2$, where F is the area of the triangle ABC.

Solution by Jiro Fukuta, Gifu-ken, Japan.

LEMMA. *Let P be an arbitrary point in ABC, with a, b, c the lengths of the sides opposite the respective vertices. Then*

$$\frac{PB}{b}\frac{PC}{c} + \frac{PC}{c}\frac{PA}{a} + \frac{PA}{a}\frac{PB}{b} \geqslant 1,$$

with equality if and only if P is the orthocenter or a vertex of ABC.

To prove the lemma, let P be the origin in the complex plane and let α, β, γ be complex numbers locating A, B, C, respectively. Then it is easy to verify the following algebraic identity

$$\frac{\beta}{(\alpha-\gamma)}\frac{\gamma}{(\alpha-\beta)} + \frac{\gamma}{(\beta-\alpha)}\frac{\alpha}{(\beta-\gamma)} + \frac{\alpha}{(\gamma-\beta)}\frac{\beta}{(\gamma-\alpha)} = 1.$$

Taking magnitudes and applying the triangle inequality yields the desired inequality. Equality holds when one of α, β, γ is 0, i.e. when P coincides with a vertex. If $\alpha\beta\gamma \neq 0$, equality holds for magnitudes if and only if each term of the equality involving α, β, γ is real, i.e., if and only if P is the orthocenter of ABC (the point of intersection of the three altitudes).

Now consider the original problem; let S be the expression on the left side of the desired inequality. The circumcircles of the triangles AB_1C_1, BC_1A_1, CA_1B_1 are concurrent in the "Miquel point," denoted by P. (Cf. Arnold Dresden, *An Invitation to Mathematics*, New York, Holt, 1935, pp. 233–239.) Let the radii of these circles and the circumcircle of ABC, respectively, be r_1, r_2, r_3, r. Using the formulas $a = 2r\sin A$, $a_1 = 2r_1\sin A$, $b_1 = 2r_2\sin B$, $c_1 = 2r_3\sin C$, we have

$$S = 8r\sin A\sin B\sin C(ar_2r_3 + br_3r_1 + cr_1r_2) = (4F/r)(ar_2r_3 + br_3r_1 + cr_1r_2),$$

because $F = 2r^2\sin A\sin B\sin C$. By the inequalities $2r_1 \geqslant PA$, $2r_2 \geqslant PB$, $2r_3 \geqslant PC$ and the lemma,

$$S \geqslant (F/r)(a\cdot PB\cdot PC + b\cdot PC\cdot PA + c\cdot PA\cdot PB) \geqslant (F/r)abc = 4F^2,$$

with the last equality from the formula $F = abc/(4r)$. Equality holds if and only if equality holds in all the inequalities used in the last step, which is true if and only if P is the orthocenter of ABC (i.e., A_1, B_1, C_1 are the feet of the three altitudes) or P coincides with one of the vertices of ABC and $A_1B_1C_1$ degenerates to one of its altitudes. Note that equality can hold for a non-degenerate triangle with obtuse angle A only when $B_1 = C_1 = A$ and A_1 is the foot of the altitude from A.

Editorial comment. The lemma in this proof appeared as long ago as T. Hayashi, "Two theorems on complex numbers," *Tohoku Math. J.* 4(1913/14) 68–70. The history of the lemma can be found in M. S. Klamkin, "Triangle inequalities from the triangle inequality," *Elem. der Mathematik*, 34(1979) 49–55.

Also solved by the proposer. Three incorrect solutions were received.

Elliptical Tangents

E 3164 [1986, 566]. *Proposed by Huseyn Demir, Middle East Technical University, Ankara, Turkey.*

Let s, t be the lengths of the tangent line segments to an ellipse from an exterior point. Find the extreme values of the ratio s/t.

Solution by Gene Arnold and Vaclav Konecny, Ferris State College, Big Rapids, MI. Consider an ellipse as the normal projection of a circle, from one plane to another in $\mathbf{R}^3$. Clearly tangents project to tangents and the minor axis of the ellipse is perpendicular to the intersection of the two planes. Since the ratio of any two intersecting tangents to the circle is 1, the extreme ratios of two such tangents to the ellipse will be attained when the ratio of one tangent to the circle to its projection is minimum while the other ratio is maximum. This happens when the projected tangents are respectively parallel to and perpendicular to the intersection of the planes. Thus the extreme ratios are those of the major axis to the minor axis, and its reciprocal.

Editorial comment. M. S. Klamkin suggested study of the more difficult problem of the extreme values of $|s - t|$.

Also solved by M. Barr (Canada), J. C. Binz (Switzerland), J. M. Cohen, J. Dou (Spain), J. Fukuta (Japan), P. L. Hon (Hong Kong), L. R. King, M. S. Klamkin (Canada), K.-W. Lau (Hong Kong), O. P. Lossers (The Netherlands), M. Pachter (South Africa), K. Schilling, J. H. Steelman, P. Tracy, D. B. Tyler, C. Vandermee (The Netherlands), and the proposer. One incorrect solution was received.

Nonunits in Commutative Rings

E 3165 [1986, 649]. *Proposed by Liviu I. Nicolaescu (student), Iassy, Romania.*

Let R be a finite commutative ring with 1, $|R| = k$. Let $N = N(R)$ denote the set of nonzero nonunits of R. Prove that $N \neq \varnothing$ forces $|N| \geqslant \lfloor\sqrt{k-1}\rfloor$.

Solution by David R. Richman, University of South Carolina, Columbia, SC. Let x be a nonzero nonunit of R, and let $A(x)$ denote the set of elements y in R such that $yx = 0$. The first isomorphism theorem shows that the additive groups Rx and $R/A(x)$ are isomorphic, and thus $|Rx| = |R|/|A(x)|$. Note also that the elements of Rx and $A(x)$ are nonunits, so $|Rx| \leqslant |N| + 1$ and $|A(x)| \leqslant |N| + 1$. Hence $(|N| + 1)^2 \geqslant |Rx|\,|A(x)| = |R| = k$. If $j = \lfloor\sqrt{k-1}\rfloor$, then $(|N| + 1)^2 > j^2$ or $|N| \geqslant j$.

Editorial comment. As mentioned by several solvers, this result was originally due to N. Ganesan, in Properties of rings with a finite number of zero divisors, *Math. Annalen*, 157(1964) 215–218). Generalizations of the original result can be found in the following:

N. Ganesan, Properties of rings with a finite number of zero divisors II, *Math. Annalen*, 161(1965) 241–246.

K. Koh, On "Properties of rings with a finite number of zero divisors," *Math. Annalen*, 171(1967) 79–80.

R. Gilmer, Zero divisors in commutative rings, this MONTHLY, 93(1986) 382–387.

D. MacHale, Wedderburn's Theorem revisited, *Bull. Irish Math. Soc.* (Sept. 1986) 44–48.

The assertion of the problem is valid without the hypothesis of commutativity.

Also solved by J. Bergen, J. M. Cohen, A. Facchini (Italy), S. Fabbri and L. Jones, R. Gilmer, R. P. Infante, I. M. Isaacs, A. A. Jagers (The Netherlands), K.-W. Lau (Hong Kong), O. P. Lossers (The Netherlands), D. MacHale (Ireland), L. E. Mattics, B. Osofsky, P. Tracy, and the proposer. One incorrect solution was received.

A Polynomial Identity

E 3166 [1986, 650]. *Proposed by Armel Mercier, University of Quebec at Chicoutimi.*

Let $n \geqslant 1$ be an integer and let f be a polynomial of degree less than or equal to n. Show the following combinatorial identity:

$$\sum_{j=1}^{n} \frac{(-1)^{j+1}\binom{n}{j} f(2j+1)}{2j(2j+1)} = \frac{n!(n+1)!2^{2n+1} f(0)}{(2n+2)!} + f'(1) - f(1) + f(1) \sum_{j=1}^{n} \frac{1}{2j}.$$

Solution I by Paul J. Zwier, Calvin College, Grand Rapids, MI. Since n is fixed, let $L(f)$ denote the left side of the desired identity. Since L is a linear functional, it suffices to show that the required identity holds for a spanning set of the real polynomials of degree $\leqslant n$.

Let $f_0(x) = x - 1$ and $f_k(x) = x(x-1)^{k-1}$ for $k \geqslant 1$. Furthermore, let $g(x) = -(1-x^2)^n$. The binomial formula yields $1 + g(x) = \sum_{j=1}^{n}(-1)^{j+1}\binom{n}{j}x^{2j}$. Applying this to the computation of L, we find $L(f_0) = \int_0^1 (1 + g(x))\,dx$, $L(f_1) = \int_0^1 \frac{1+g(x)}{x}\,dx$, and $L(f_k) = [(xD)^{k-2}(1+g(x))]_{x=1}$ for $k \geqslant 2$, where xD is the differentiation operator followed by multiplication by x (actually, the formula for f_k holds also for f_1.)

Now we evaluate these expressions using $g(x) = -(1-x^2)^n$. For f_0, we substitute $x = \sin z$ to get $L(f_0) = 1 - \int_0^{\pi/2}\cos^{2n+1}z\,dz = 1 - [n!(n+1)!2^{2n+1}/(2n+2)!]$. For f_1, we let $u = 1 - x^2$ to obtain $L(f_1) = (1/2)\int_0^1(1-u^n)\,du/(1-u) = \sum_{j=1}^{n} 1/(2j)$. Since $g(1) = 0$, $L(f_2) = 1$. Finally, since $(1-x^2)^{n-m}$ divides $(xD)^m(1+g(x))$ for $1 \leqslant m \leqslant n-1$, we have $L(f_k) = 0$ for $3 \leqslant k \leqslant n$.

We have chosen the spanning set f_k so that $f_k(0) = -\delta_{k0}$, $f_k(1) = \delta_{k1}$, and $f_k'(1)$ equals 1 for $0 \leqslant k \leqslant 2$ and 0 for $3 \leqslant k \leqslant n$. Here δ denotes the Kronecker delta. By examination, then, the values computed for $L(f_k)$ above are the values of the right side for f_k. Hence each real polynomial of degree at most n satisfies the identity.

Solution II by Donald E. Knuth, Stanford University, Stanford, CA. The identity holds as well when f has degree $n+1$. Let f be an arbitrary polynomial of degree at most $n+1$. We can write f in terms of a polynomial g of degree less than n as $f(2j+1) = a + (2j+1)(b + 2jg(j))$, for some constants a and b. The given sum is then

$$a\sum_{j>0}(-1)^{j+1}\binom{n}{j}\left(\frac{1}{2j} - \frac{1}{2j+1}\right) + b\sum_{j>0}(-1)^{j+1}\binom{n}{j}\frac{1}{2j}$$
$$+ \sum_{j>0}(-1)^{j+1}\binom{n}{j}g(j)$$
$$= \frac{a+b}{2}\sum_{j>0}(-1)^{j+1}\binom{n}{j}\frac{1}{j} + \frac{a}{2}\left(-2 + \sum_{j\geqslant 0}(-1)^{j}\binom{n}{j}\frac{1}{j+1/2}\right) + g(0),$$

since $\sum_{j\geqslant 0}(-1)^{j+1}\binom{n}{j}g(j) = 0$ is the nth difference of g at 0. The second sum

evaluates to $2^{2n+1}n!^2/(2n+1)!$ by setting $x = 1/2$ in the partial fraction expansion

$$\sum_{j\geqslant 0}(-1)^j\binom{n}{j}\frac{1}{j+x} = \frac{n!}{x(x+1)\cdots(x+n)}. \qquad (*)$$

The first sum evaluates to $\sum_{j=1}^{n}1/j$ by subtracting $1/x$ from both sides of $(*)$, letting $x \to 0$, and applying l'Hospital's rule. Finally, we use the facts $f(0) = a$, $f(1) = a + b$, $f'(1) = b + g(0)$ to obtain the desired right side.

Editorial comment. Most solvers showed that the identity holds for the powers x^k, which involves a somewhat lengthier calculation than that of Solution I above. Michiel H. M. Smid used induction on n. The fact that the identity also holds for degree $n+1$ can be obtained from Solution I by using the remark that $(1-x^2)$ divides $(xD)^{n-1}(1+g(x))$ to conclude that $L(f_{n+1}) = 0$.

Also solved by J. C. Binz (Switzerland), L. Kuipers (Switzerland), K.-W. Lau (Hong Kong), O. P. Lossers (The Netherlands), M. H. M. Smid (The Netherlands), D. B. Tyler, University of South Alabama Problem Group, Western Maryland College Problems Group, and the proposer.

Continuous and Differentiable

E 3173 [1986, 733]. *Proposed by H. W. Oliver, Williamstown, MA.*

Let the real-valued function f be differentiable on the open interval (a, b). For $(x, y) \in D = \{x, y): a < x < b,\ a < y < b\}$, define

$$F(x, y) = \frac{f(x) - f(y)}{x - y} \quad \text{for } x \neq y,$$
$$F(x, x) = f'(x).$$

Give a necessary and sufficient condition (in terms of the function f) that

(i) F be differentiable at each point of D;

(ii) F be continuously differentiable on D.

Solution by O. P. Lossers, Eindhoven University of Technology, The Netherlands.

(i) The condition is differentiability of f'. We use the interpretation of the derivative as a linear approximation to a function. I.e., a function g is differentiable with derivative $g'(x)$ at x if and only if $g(x+h) = g(x) + g'(x)h + o(h)$, where $o(h)$ denotes a function whose ratio with $|h|$ goes to 0 as h goes to 0. Similarly, for two variables,

$$F(x+h, y+k) = F(x, y) + \big(D_1F(x, y), D_2F(x, y)\big)\cdot(h, k) + o(|(h, k)|).$$

If F is differentiable at each point of D, then clearly f' is differentiable on (a, b). Conversely, if f is twice differentiable, then F is differentiable. This is immediate for off-diagonal points; to prove that F is differentiable at the point (x, x), we consider the function $g(t)$ defined by

$$g(t) = f(x+t) - tf'(x) - \frac{1}{2}t^2f''(x).$$

Since $g'(0) = 0$ and $g''(0) = 0$, we have $g'(t) = o(t)$. Applying the mean value theorem to the interval between h and k, we obtain a value c between h and k such

that

$$\frac{g(h)-g(k)}{h-k}=g'(c)=o(|h|+|k|).$$

The expression in terms of g on the left equals $F(x+h, x+k)-F(x,x)-(1/2)(h+k)f''(x)$. Since $o(|h|+|k|)=o(|(h,k)|)$, this proves that F is differentiable at (x,x) with derivatives $D_1F(x,x)=D_2F(x,x)=f''(x)/2$.

(ii) The condition is continuous differentiability of f'. If F is continuously differentiable, then f'' is continuous. Conversely, if f'' is continuous, we claim that D_1F and D_2F are continuous. For all $(x,y)\in D$, integration by parts yields $\int_0^1 f'(y+t(x-y))\,dt=F(x,y)$. Hence

$$D_1F(x,y)=\int_0^1 tf''(y+t(x-y))\,dt,$$

which is clearly a continuous function.

Also solved by D. Doster, K. Schilling, and the proposer.

Convergent and Divergent Series

E 3174 [1986, 733]. *Proposed by Albert Wilansky, Lehigh University, Bethlehem, PA.*

Let $a\in cs$ mean that Σa_n converges. Show that

(i) there exists $b\notin cs$ such that $\inf_{a\in cs}\sup_n\left|1-\dfrac{a_n}{b_n}\right|=0$,

and

(ii) there exists $a\in cs$ such that $\inf_{b\notin cs}\sup_n\left|1-\dfrac{a_n}{b_n}\right|=0$,

Editorial comment. The problem tacitly assumes that $b_n\neq 0$ for all n.

Solution by W. Hensgen, Universität Regensburg. Both assertions are implied by the existence of $a\in cs$ and $b\notin cs$ such that $a_n, b_n\neq 0$ and $a_n/b_n\to 1$. In case (i) we can take

$$b_n=(-1)^n\{1+(-1)^n/\log(n+1)\}/n \qquad (n=1,2,\dots)$$

and, for given k,

$$a_n^{(k)}=b_n \quad\text{for}\quad 1\leqslant n\leqslant k, \qquad a_n^{(k)}=(-1)^n/n \quad\text{for}\quad n\geqslant k.$$

An analogous example can be given for (ii).

Also solved by D. Callan, J. N. Fitch, E. Hertz, O. P. Lossers (Netherlands), R. Neidinger, M. J. Reed, A. Riese, A. V. Stanoyevich, A. Stenger, L. Thurston, D. M. Wells, and the Univ. of South Alabama Problem Group, and partially by the proposer.

A Sined Matrix

E 3178 [1986, 812]. *Proposed by G. A. Hively, Lockheed Palo Alto Research Laboratory, CA.*

Let x_j and y_k be arbitrary real numbers. Show that the $n \times n$ matrix

$$S = \left(\sin(x_j + y_k)\right)$$

is singular if $n \geqslant 3$.

Solution I by Jack Tomsky, Lockheed Palo Alto Research Laboratory, CA. Let A be the $n \times 2$ matrix whose jth row is $(\sin x_j, \cos x_j)$, and let B be the $n \times 2$ matrix whose kth row is $(\cos y_k, \sin y_k)$. Because $\sin(x_j + y_k) = \sin x_j \cos y_k + \cos x_j \sin y_k$, we have $S = AB^T$. Thus S has rank at most 2 and is singular for $n \geqslant 3$.

Solution II by Bruce Dearden, University of North Dakota, Grand Forks, ND. The matrix S is the imaginary part of the matrix $E = (\exp i(x_j + y_k))$. Let v and w be the column vectors whose jth entries are e^{ix_j} and e^{iy_j}, respectively. Since for $n \geqslant 3$ the entries of v are linearly dependent over the reals, the equation $E = vw^T$ implies that the rows of S are linearly dependent over the reals.

Editorial comment. Jesús Ferrer (Spain) and Pei-Yuan Wu (Taiwan) both noted that the special case with $x_j = y_j$ for all j appears as Problem 294 in Faddeev and Sominskii, *Problems in Higher Algebra*. A number of readers pointed out that analogous propositions are true when sin is replaced by other functions with suitable addition formulas.

Also solved by the proposer and 40 other readers.

ADVANCED PROBLEMS

6576. *Proposed by Hans U. Gerber, University of Lausanne, Switzerland.*

Suppose $X_1, X_2, \ldots$ are independent identically distributed real random variables with $E(X_k) = \mu$. Put

$$S_k = X_1 + X_2 + \cdots + X_k \quad \text{for} \quad k = 1, 2, \ldots .$$

(a) If $\rho < \mu < 1$, where $\rho = -0.278465 \cdots$ is the real root of $xe^{1-x} = -1$, show that the series

$$\sum_{k=1}^{\infty} S_k^k e^{-S_k}/k!$$

converges with probability one.

(b) If $X_1, X_2, \ldots$ are positive and if $\mu < 1$, show that the expectation of

$$\sum_{k=1}^{\infty} S_k^k e^{-S_k}/k!$$

is $\mu/(1-\mu)$.

(c)* In (b) is it possible to relax the condition that the random variables are positive? For example, would it suffice to assume $E(|X_k|) < \infty$ and $\rho < \mu < 1$?

6577. *Proposed by B. Bagchi, G. Misra, and N. S. N. Sastry, Indian Statistical Institute, Calcutta, India.*

(a) Let H be an infinite-dimensional inner-product space. Suppose that finitely many closed balls cover the surface S of the unit ball B in H. Prove that these balls also cover the center of B.

(b) Does the above assertion remain valid if H is any infinite-dimensional, normed linear space?

6578. *Proposed by S. T. Stefanov, Sofia, Bulgaria.*

A fluid flows on a 2-sphere S with continuous velocity vector $\bar{v}(x, t)$. Let $T(x, t)$ be the temperature of the fluid at the point $x \in S$ at the time t. Assume that $T(x, 0) < T_0$ and $T(x, 1) > T_0$ for all $x \in S$ and some T_0. Prove that there exist $x_0 \in S$ and $t_0 \in (0, 1)$ such that $T(x_0, t_0) = T_0$ and $\bar{v}(x_0, t_0) = 0$.

SOLUTIONS OF ADVANCED PROBLEMS

Critical Points of Gateaux Differentiable Functions

6467 [1984, 441]. *Proposed by Bu Qi-yue (student), Shanghai Jiao Tong University, People's Republic of China.*

Suppose that g: $D \subset R^n \to R^1$ has a second Gateaux derivative at $X_0 \in \text{int } D$. Suppose also that X_0 is a critical point of g and $g''(X_0) > 0$. Must X_0 be a proper local minimizer of g? Answer the same question with "Fréchet" instead of "Gateaux."

Editorial Remark. A function ϕ: $D \to R$, where D is an open set in R^n, is said to be Gateaux differentiable at a point $x_0 \in D$ if there exists a linear function

$$\phi'_{G, x_0}: R^n \to R$$

such that

$$\phi'_{G, x_0}(h) = \lim_{t \to 0} \frac{\phi(x_0 + th) - \phi(x_0)}{t}, \quad \text{all } h \in R^n.$$

The function ϕ is said to be Fréchet differentiable at x_0 if there exists a linear function

$$\phi'_{F, x_0}: R^n \to R$$

such that

$$\lim_{\|h\| \to 0} \frac{\phi(x_0 + h) - \phi(x_0) - \phi'_{F, x_0}(h)}{\|h\|} = 0, \quad \text{all } h \in R^n.$$

The functions ϕ'_{G, x_0} and ϕ'_{F, x_0}, if they exist, are unique and are called, respectively, the Gateaux and the Fréchet derivatives of ϕ at x_0.

Solution by the proposer. It is clear that if g is twice Fréchet differentiable at X_0, then the stated conditions guarantee a proper local minimizer at X_0. But if g is assumed only to be twice Gateaux differentiable at X_0, then the situation is different.

For example, fix $n = 2$ and let D be the unit disc in the plane. For $0 \leq \theta < 2\pi$ let C_1 and C_2 be the curves

$$r = \cos(\theta/4), \qquad r = (1/2)\cos(\theta/4),$$

respectively. Denote by S the open set consisting of the points between the two curves (thus S is a band that narrows while spiraling into the origin). Let

$$g(x) = \begin{cases} \|x\|^2 & x \notin S \\ -1 & x \in S. \end{cases}$$

For h, $k \in D$, a computation shows that

$$g'(0,0) = 0$$

and

$$h \cdot g''(0,0)k = 2h \cdot k.$$

Therefore, $g''(0,0)$ is positive definite, but obviously 0 is not a local minimizer of g.

No other correct solutions were received.

Lipschitz vs. Hölder

6529 [1986, 743]. *Proposed by O. Hájek, Case Western Reserve University.*

Let $f_1, f_2, \ldots$ be a sequence of functions from R^1 to R^1 with common Lipschitz constant λ. Show that each continuous function f with

$$(\text{graph } f) \subset \bigcup_{k=1}^{\infty} (\text{graph } f_k)$$

also has Lipschitz constant λ. (The condition on the graph of f may be restated as follows: for each real x there is a k with $f(x) = f_k(x)$.) Is this true if the Lipschitz condition is replaced by a Hölder condition of order α for some fixed α with $0 < \alpha < 1$?

Combined solution by the proposer and M. J. Pelling, London, England. By considering $\lambda x \pm f_k(x)$ one need only prove this: if all f_k are increasing, then so is each continuous f with

$$(\text{graph } f) \subset \bigcup (\text{graph } f_k).$$

Assume the contrary, that somewhere $f(a) > f(b)$ with $a < b$. We shall show that f strictly decreases on an uncountable set, and a contradiction will ensue. By decreasing the interval $[a, b]$ if necessary, we may even assume that $f(a) \geqslant f(x) \geqslant f(b)$ for all $x \in [a, b]$. By continuity, for each $t \in [f(b), f(a)]$, we may define

$$g(t) := \min\{x \in [a, b]: f(x) = t\}.$$

Then $f(g(t)) = t$, and $t < s$ implies $g(t) > g(s)$ (since, by the intermediate value theorem, both $=$ and $<$ lead to a contradiction).

For each $t \in [f(b), f(a)]$ there is an index k such that

$$t = f(g(t)) = f_k(g(t)).$$

To different t's there must correspond distinct k's. Indeed, if $t < s$ and $s =$

$f_m(g(s))$, then $g(t) > g(s)$; since $t < s$ and f_k is increasing, we cannot have $m = k$. Thus there are continuum-many t's but only countably many k's: a contradiction.

The Lipschitz condition cannot be replaced by a Hölder condition. For example, if $\alpha = 1/2$ and $\lambda = 1$ define

$$f_1(x) = \sqrt{x}, \qquad 0 \le x < \infty,$$

$$f_2(x) = \begin{cases} 1 & 0 \le x \le 1, \\ 1 + \sqrt{x-1} & 1 \le x < \infty, \end{cases}$$

and

$$f(x) = \begin{cases} f_1(x) & 0 \le x \le 1, \\ f_2(x) & 1 \le x < \infty. \end{cases}$$

Then f_1 and f_2 satisfy the order 1/2 Hölder condition with constant $\lambda = 1$, but f does not since

$$|f(2) - f(0)| = 2 > 1(2-0)^{1/2} = \sqrt{2}.$$

Here the counterexample is due to M. J. Pelling, while the affirmative part is due to the proposer.

Also solved by Pavol Ralbovský (Czechoslovakia).

Graphs with Many Horizontal Chords

6530 [1986, 743]. *Proposed by Wolfgang Walter, Universität Karlsruhe, Federal Republic of Germany.*

Let f be a real-valued continuous function on $(0,1)$ and $\{a_n\}, \{b_n\}$ two sequences of real numbers such that $a_n \ne b_n$, $n = 1, 2, 3, \ldots$, and

$$\lim a_n = \lim b_n = 0.$$

Assume that for each x in $(0,1)$ there exists an $N = N(x)$ for which

$$f(x + a_n) = f(x + b_n), \qquad n \geqslant N(x).$$

Must f be a constant function if $0 < b_n < a_n$?

*What if $b_n = -a_n$?

Partial solution by O. P. Lossers, Eindhoven University of Technology, The Netherlands. The answer to the first question is "No." This is seen as follows. Write $x \in (0,1)$ in decimal notation,

$$x = \sum_{n=1}^{\infty} \varepsilon_n . 10^{-n} \qquad (0 \le \varepsilon_n \le 9).$$

For $x \in (0,1)$ let $M(x)$ be the smallest integer n such that $\varepsilon_{2n+1} \ne 0$. Then define

$$f(x) := \sum_{n=1}^{M(x)} \varepsilon_{2n} . 10^{-n} + 10^{-M(x)}.$$

(If $\varepsilon_{2n+1} = 0$ for all n then define $f(x) := \sum_{n=1}^{\infty} \varepsilon_{2n} . 10^{-n}$). It is easily verified that f is uniquely determined. Let $a_n = 2 \cdot 10^{-2n-1}$ and $b_n = 10^{-2n-1}$. Then f, a_n, b_n satisfy the given conditions but f is not constant. The function f is continuous,

because for all n

$$|x-y| \le 10^{-2n} \Rightarrow |f(x)-f(y)| \le 10^{-n}.$$

Next, if $M(x)=\infty$ then clearly $f(x+a_n)=f(x+b_n)$. Finally, if $M(x)=M$, where x is represented so that it does not end in a string of 9's and N is the first position after M with $\varepsilon_N < 9$, then $f(x+a_n)=f(x+b_n)$ for $n > N/2$.

Editorial comment. Adam Riese independently provided an essentially identical solution, and also showed that if $b_{n+1}=a_n$, then f must be constant. His argument is to assume that

$$f(x_1) \neq f(x_2), \qquad 0 < x_1 < x_2 < 1,$$

and define

$$C = \{x \in [x_1, x_2]: f(x) = f(x_1)\}.$$

Then C is closed and hence has a largest number c; clearly $f(x) \neq f(c)$ for x in $(c, x_2]$. Now

$$f(c+a_n) = f(c+b_n) = f(c+a_{n-1})$$

for n sufficiently large. Since

$$\lim_{n\to\infty} f(c+a_n) = f(c),$$

it follows that $f(c+a_n)=f(c)$ for n sufficiently large, a contradiction for $0 < a_n$.

M. J. Pelling and John Henry Steelman both pointed out the similarity of their counterexample functions to the classical Cantor singular function. Steelman observed that if $f(x+a_n)=f(x-a_n)$ for large n and

$$S = \{x \in (0,1): f \text{ is constant in a neighborhood of } x\},$$

then S is an open dense subset of $(0,1)$. This follows by a Baire category argument. If f is not constant, the complement of S must be uncountable. This suggests that any such nonconstant f must "resemble" a Cantor function.

No other solutions were received.

Holomorphic Functions Satisfying a Triangle Equality

6533 [1987, 81]. *Proposed by Boo Rim Choe (graduate student), University of Wisconsin-Madison.*

Let t be a fixed real number. Find all functions $f(z)$ holomorphic in a disc D centered at $z=0$ that satisfy

$$|f(z)| = |f(x)| + t|f(y)|$$

for all $z = x+iy$ in D.

Solution by O. P. Lossers, Eindhoven University of Technology, The Netherlands. Aside from $f(z)\equiv 0$, the only possible solutions are $f(z)=Ke^{\rho z}$ for real ρ and complex K, and $f(z)=Az^2$ for complex A. These can only occur when $t=0$ and $t=1$, respectively. To prove this it is natural to consider the cases $f(0)\neq 0$ and $f(0)=0$ separately. In the first case t must be zero, so $|f(z)|$ depends only on x,

i.e.,

$$f(z) = g(x)\, e^{ih(x, y)},$$

where g and h are real, twice-continuously differentiable functions of their variables. The Cauchy-Riemann equations, applied either to $f(z)$ or to $\log f(z)$, quickly yield

$$g'(x) = g(x)\frac{\partial h}{\partial y}, \quad \text{and} \quad 0 = g(x)\frac{\partial h}{\partial x}.$$

Hence h depends only on y and

$$\frac{g'(x)}{g(x)} = h'(y) = \rho$$

for some real constant ρ. These differential equations are easily solved to yield $f(z) = c_1 e^{ic_2} e^{\rho z}$. The case $f(0) = 0$ but $f(z) \not\equiv 0$ is a bit harder. Say

$$f(z) = Az^k + \sum_{m=k+1}^{\infty} c_m z^m, \qquad A \neq 0$$

is the power series of f for $z = re^{i\phi}$ in a small neighborhood of zero. Then

$$\left|\frac{f(z)}{z^k}\right| = \left|\frac{f(x)}{x^k}\frac{x^k}{z^k}\right| + t\left|\frac{f(y)}{y^k}\frac{y^k}{z^k}\right| \qquad (*)$$

yields

$$|A| = |A|\big(|\cos\phi|^k + t|\sin\phi|^k\big), \quad \text{all } \phi,$$

so $t = 1$ and $k = 2$. Now any solution other than $f(z) = Az^2$ would lead to a solution of the form

$$f(z) = z^2 + Be^{i\theta}z^l + O(z^{l+1}),$$

where $B > 0$ and $l \geqslant 3$. Then $r \to 0$ together with $(*)$ for $k = 2$ implies

$$|1 + Be^{i\theta}r^{l-2}e^{(l-2)i\phi}| + O(r^{l-1}) = \cos^2\phi\,|1 + Be^{i\theta}r^{l-2}\cos^{l-2}\phi| + \sin^2\phi\,|1 + Be^{i\theta}r^{l-2}\sin^{l-2}\phi|.$$

Since

$$|1 + \rho e^{i\alpha}| - 1 = \rho\cos\alpha + O(\rho^2)$$

for small ρ, and $l \geqslant 3$ implies $2l - 4 \geqslant l - 1$, the difference of the left- and right-hand sides of the equation in the previous sentence is

$$Br^{l-2}\big[\cos(\theta + (l-2)\phi) - \cos\theta\cos^l\phi - \cos\theta\sin^l\phi\big] + O(r^{l-1}).$$

From this we first deduce

$$\cos(\theta + (l-2)\phi) = \cos\theta\cos^l\phi + \cos\theta\sin^l\phi,$$

and then (differentiate and set $\phi = 0$ to see that $\theta = n\pi$, n an integer) the absurdity

$$\cos(l-2)\phi = \cos^l\phi + \sin^l\phi, \quad \text{all } \phi.$$

Also solved by Rainer Brück (Federal Republic of Germany), B. Crofoot and K. Kearnes (jointly), Kee-wai Lau (Hong Kong), Douglas B. Tyler, and the proposer. Two incomplete solutions were received.

REVIEWS

EDITED BY JOSEPH KONHAUSER
Department of Mathematics, Macalester College, St. Paul, MN 55105

Metric Spaces: Iteration and Application. By Victor Bryant. Cambridge University Press, Cambridge, 1985. v + 105 pp.

MICHAEL EDELSTEIN
Department of Mathematics, Dalhousie University, Halifax, Nova Scotia B3H 3J5 Canada

Classes in mathematics, made up of students of widely varying backgrounds, pose something of a challenge to an instructor. To reach a diverse audience, including the less capable individuals in it, a teacher must show that the material is both interesting and useful. Thus extra time for illustrations, examples and applications is needed; furthermore, a slower lecture pace must be adopted. In these circumstances it may be difficult to cover adequately the anticipated course content. In general, the experienced teacher knows how to strike an appropriate balance. In the book under review such a balance is, unfortunately, missing.

Designed as a course for combined honors students, the book is meant to be elementary or—as the author puts it—"down to earth." To make the material easier to absorb much of the standard metric space theory is either omitted or diluted. (For example, separability, total boundedness, the process of completion are omitted; connectedness is dealt with all too briefly.) Motivation, on the other hand, is emphasized throughout.

The starting point and recurrent theme, taking up the greater part of the book, is that of iteration; i.e., sequences arising from iterative procedures. Here a wealth of interesting examples is given, culminating in the proof of an implicit function theorem. Unquestionably, this topic is presented in an appealing manner; and in a much larger book it would have been entirely appropriate. Some paring down of this material seems to be called for, and it may be of some interest to consider the following scheme for a more condensed, yet forceful, treatment of this particular topic.

Before plunging into course work proper, students might be challenged with a simple concrete exercise. For example, they could be asked to produce a program (in Basic or any other computer language) solving, to a reasonable approximation, the equation $x - \cos x = 0$ $(0 \leq x \leq 1)$. (With the current popularity of computer science courses this should be quite easy.) Satisfied that this was accomplished, the instructor might ask for the seemingly easier problem of solving the same equation with sine replacing cosine, but with initial approximation $x_0 = 1$. Students will, of course, unavoidably find, that for the former equation, twenty iterations suffice to produce the approximation $x \approx 0.739$, correct to the last digit; while, in contrast, even after 10,000 iterations, $x \approx 0.017$ is incorrect in the last two digits.

Why should this be so? Do one or both of the above two sequences converge? If both do, why is there such a disparity in the rate of convergence? In particular, why is it so much harder to attain an improved approximation for the second equation (when starting with $x_0 \neq 0$)?

These are thought-provoking questions that lead quite naturally to a discussion of contractions and other mappings with related properties. A more comprehensive study of such problems might have to be postponed or omitted altogether, but with a little preparation, it can be pointed out in this context that iterates under a contraction always form a Cauchy sequence.

The above scheme, if implemented, would still leave a text, already rather shorthanded in theory, even more lacking. The need for its enrichment would thus be more compelling. What additional material should be included? There is an embarassment of riches here and any practical answer must be somewhat arbitrary.

At a minimum one would want to see precompactness included so as to intertwine, via that notion, compactness and completeness. In the three C's mentioned by the author, connectedness should be restored to its rightful place alongside completeness and compactness, displacing closedness(!). And, of course, connectedness has to be given more prominence. Separability and second countability properties could be included, with the intention of exhibiting the delightful embedding of such spaces into the Hilbert cube.

Finally, a remark or two about general comments. A class filled with budding chemists, engineers, etc., should be issued what may be called an invitation to mathematics. By this I mean the general, historical, and cultural asides which show that mathematics is an evolving human endeavor. More specifically, the genesis of the notion of a metric space and its position in various branches of mathematics should be treated. Further, to cultivate and pay homage to the human aspect of mathematical creativity, names of outstanding contributors to the discipline should be mentioned. A course in metric spaces looks arid and deprived without the names of Frèchet, Hausdorff, Urysohn, and Banach in it.

Mathematics for Liberal Arts: A Problem Solving Approach, by Rick Billstein and Johnny W. Lott. The Benjamin/Cummings Publishing Co., 1986, xvii + 646 pp.

Mathematical Ideas, fifth edition, by Charles D. Miller and Vern E. Heeren, Scott, Foresman and Company, 1986, xiv + 751 pp.

Alvin M. White
Department of Mathematics, Harvey Mudd College, Claremont, CA 91711

Billstein and Lott start each chapter with a preliminary problem from everyday experience which motivates the topics of the chapter. After definitions, techniques, examples and exercises, the chapter ends with an analysis of the solution of the preliminary problem following Pólya's advice: Understanding the problem; Devising a plan; Carrying out the plan; Looking back. Scattered through the book are brain teasers whose statements are easy to understand but whose solutions are not obvious or easy. The brain teasers would be appropriate challenges for groups of students to ponder.

Miller and Heeren is the fifth edition of a book that is more straightforward in its presentation of definitions, techniques, examples and exercises.

The first book uses cartoons "to add a lighter touch." The second book features postage stamps of famous mathematicians in the margin with a postage stamp's worth of explanation of why that mathematician is so honored.

Both books are designed for liberal arts students, and in that genre they are typical. Both offer essentially the same content at the same level in similar styles. The topics are standard, elementary finite mathematics. Almost all of the material is algorithmic: Here is a method—There is a set of exercises that can be done using that method. Except for Babylonian and Hindu-Arabic numeration, the material is presented as a timeless collection of facts and techniques with no history or evolution. The historical notes in one and the postage stamps of mathematicians in the other are hardly sufficient to arouse the curiosity of the reader except to think, "Who are these people and why are their pictures in the margin of my book?" The many separate topics have no evident connection, very little motivation, and no hint or guide to possible understanding at a deeper level.

Most mathematics for liberal arts texts miss an opportunity to serve the liberal arts students. What is usually offered is elementary mathematics, superficially treated, presented with a finality that forecloses further inquiry, connections or development.

An alternative approach could put mathematics in the center of the liberal arts. Instead of formulas and routines explained in a few pages, if at all, ideas would receive the main attention. Certain ideas have made mathematics possible and have enriched the liberal arts. Mathematics, as intellectual history, has changed our world view and altered our modes of thought. The traditional text introduces Cartesian coordinates in order to present the equations of a straight line and a circle, followed by exercises. The equation of a straight line and the location of its x and y intercepts are of little concern to liberal arts students. The enormous influence of Descartes in mathematics as well as in philosophy throughout Europe ([**3**], [**8**]), however, is completely ignored. An opportunity to connect mathematics with liberal arts is lost.

A valid question is whether poorly understood technicalities are more valuable for liberal arts students than important concepts and discoveries of mathematics. If straight lines are introduced, they could be followed by the idea of parallel lines and a discussion of Euclid's fifth postulate. A valuable contribution to liberal arts from mathematics would flow from a discussion of axiomatic systems. The story of the search for a proof of the fifth postulate is an exciting chapter of mathematics as a human adventure. The discovery of Non-Euclidean geometry, which resolved the search, has been compared to the Copernican revolution. Here are mathematical adventures and liberal arts combined.

Relevance to Liberal Arts. Liberal arts students are interested in their place in the world and a sense of how they got there. The principles underlying our world view are intimately related to mathematics at least since Pythagoras. The role of mathematics in modern science is relevant, and the place of modern science in coloring our ways of thought and mentality is fundamental. Writing about the rise of modern science, Whitehead [**8**] declares, "Since a babe was born in a manger, it may be doubted whether so great a thing has happened with so little stir... The new mentality is more important even than the new science and the new technology. It has altered the metaphysical presuppositions and the imaginative content of our minds, so that now the old stimuli provoke a new response."

Mathematics is more than facts to be memorized or techniques to be mastered. Most texts, however, give us only facts and techniques. S. K. Stein described that

state of affairs in his talk, "Gresham's Law: Algorithm Drives out Thought" [6]. "Gresham's law in economics says, 'Bad money drives good money out of circulation...' Gresham's law in mathematical pedagogy can be stated—'Algorithm drives out thought'. 'The robotic displaces humanistic'. 'Cultivation of algorithms replaces concern for thinking and writing'... Algorithms, of course, are good and must be taught.... But the temptation to emphasize drill over understanding is almost irresistible"—and algorithms are easier to teach and test.

Imagination and Invention. Albert Einstein is quoted on posters, tee shirts and bumper strips:

Imagination is More Important Than Knowledge.

That sentiment should be evident in mathematics, especially for liberal arts students. Imagination and invention are what make mathematics part of the liberal arts. Whitehead [8] describes Pure Mathematics as the most "original creation of the human spirit." But our students see neither creation nor spirit. Regarding the basic topics that we teach, the question is never raised, 'Why so?' or 'How does one arrive at them?' "Yet all these matters must at one time have been goals of an urgent quest, answers to burning questions, at the time, namely, when they were created. If we were to go back to the origins of these ideas, they would lose that dead appearance of cut-and-dried facts and instead take on fresh and vibrant life again" [7].

Thematic Continuity. The value of mathematics to liberal arts students is not in a few facts and techniques. The statement and proof of the Pythagorean theorem are not as important as the *idea* of proof. The invention and development of deductive reasoning is one of the triumphs of intellect. What is the purpose of proof in mathematics and how does it compare with proof in physics or anthropology? These are questions that are in the mainstream of liberal arts—certainly more so than Venn diagrams or BASIC programming.

The Pythagoreans offer the student much more than a famous theorem. Toeplitz [7] relates two proofs of the irrationality of $\sqrt{2}$ that he attributes to the Pythagoreans. The second proof uses an elementary geometrical construction instead of facts about even and odd numbers, and proves that it is impossible to find a common measure for the side and diagonal of a square. The meaning of an impossibility proof is certainly worthy of discussion. Toeplitz writes of this first impossibility proof. "This great discovery, more than anything else, inaugurated the character of modern mathematics."

Two thousand years later Fermat uses the same method, now called "infinite descent," to prove several theorems, including a simple case of his last theorem. Thematic continuity is evident in mathematics, science and liberal arts.

Almost three hundred years after Fermat, in his Nobel Prize speech in 1922 Niels Bohr describes how the concept of the quantum had led gradually to a systematic classification of any electron in an atom, offering a complete explanation of the remarkable relationships between the physical and chemical properties of the elements, as expressed in the periodic table of Mendeleev. "Such an interpretation of the properties of matter appeared as a realization; even surpassing the dreams of the Pythagoreans of the ancient ideal of reducing the formulation of the laws of nature to considerations of pure number." [1]

Examples. Using appropriate mathematics as illustrations, texts and teachers can trace the development of vague ideas to polished concepts. The discovery, evolution and meaning of rigor is an example that might show that mathematics involves more than memorization and might indicate some reasons why mathematics has endured so vigorously. The role of human needs and invention might become evident with a description of Zeno's paradox and a discussion of the infinite from Pythagoras to Cantor to Hilbert. A discussion about the paradoxes and other crises of set theory would contribute to appreciation of mathematics as a human activity. Technical definitions and routines should include meaning and context. Liberal arts students are not automatons who can proceed without understanding.

It would be worth introducing some elementary calculus so that the class could divide into two debate teams and a jury to consider Bishop Berkeley's attack on Newton's ideas on one side versus the ideas of Newton's defenders on the other. The debate-discussion would clarify the issues and ideas and be in the spirit of liberal arts.

Influence of Newton. Should the formulas for derivatives be the extent of the liberal arts students' acquaintance with Newton? E. A. Burtt describes Newton as the one Englishman whose authority and influence in modern times has rivaled that of Aristotle over the late medieval epoch. Newton was even more honored and respected in the eighteenth century than he is today. He not only brought glory to his country, "but has done honor to human nature by having extended the greatest and most noble of our faculties, reason, to subjects, which till he attempted them, appeared to be wholly beyond the reach of our limited capacities." Newton set the scientific program for two hundred years. The whole orientation of philosophy was shifted by his ideas [**3**]. It would seem unfair, especially to liberal arts students, to represent Newton's contributions by formulas that can now be produced by calculator keys.

Whitehead reviews the genesis of modern science in the Century of Genius [**8**]. "The subject of the formation of the three laws of motion and of the law of gravitation deserves critical attention. The whole development of thought occupied exactly two generations. It commenced with Galileo and ended with Newton's *Principia*; and Newton was born in the year Galileo died. Also the lives of Descartes and Huyghens fall within the period occupied by these great terminal figures. The issue of the combined labours of these four men has some right to be considered as the greatest single intellectual success which mankind has achieved..." The law of inertia "contains the repudiation of a belief which had blocked the progress of physics for two thousand years. It also deals with a fundamental concept which is essential to scientific theory; I mean, the concept of an ideally isolated system. This conception embodies a fundamental character of things, without which science or indeed any knowledge on the part of a finite intellect would be impossible."

Those are heady sentiments that are ignored by many authors of mathematics for liberal arts texts. Although mathematics is central to modern science and modern western culture, most texts bypass these important aspects of mathematics and offer, instead, essentially no intellectual content or challenge. The mind is certainly not stretched or inspired by the contents of these texts. Is math or science literacy better enhanced by memorizing algorithms without understanding, or by considering how certain mathematical and scientific ideas changed the direction of civiliza-

tion? Of course, it may be necessary to study preliminary ideas and concepts in order to understand the pivotal moments. But then the preliminary study is given a purpose and direction.

Report of MSEB. It is difficult to break with tradition, as I advocate. But the tradition has not served us well. *SIAM News* (May 1987) reports on the conclusions of the Mathematical Sciences Education Board (MSEB). The mathematical skills of American precollege students, when compared to comparable students in other industrialized countries, rank near the bottom. "Boring and repetitious curriculum, low expectations, and negative attitudes have been identified as leading culprits." Other disappointing conclusions are: "An average European or Japanese student is typically better prepared than the average American; U.S. standardized tests measure largely irrelevant skills; U.S. students spend about 40 hours a year more than Japanese students on mathematics; and the U.S. curriculum's focus on computation is one of its biggest weaknesses."

The Project Physics Course. If breaking with tradition for texts is considered radical or just avant-garde among mathematicians, it is old hat among physicists. While mathematicians were creating and refining the New Math, physicists were creating and publishing The Project Physics Course in high school and college versions. In 1952 Gerald Holton wrote *Introduction to Concepts and Theories in Physical Science* (Addison-Wesley). This was followed by *Foundations of Modern Physical Science* in 1958 (with Duane H. D. Roller) and the second edition of Concepts and Theories (with Stephen Brush) in 1972. In 1970 Holt, Rinehart and Winston published *The Project Physics Course* which was an integrated combination of text, laboratory experiment films, transparencies, and teacher resource books.

The following quotes are from the prefaces of *Concepts and Theories*. [5]

> Perhaps the most striking characteristic of this book is the emphasis it puts on the nature of discovery, reasoning, and concept formation in science as a fascinating topic in its own right. This means that the historical and philosophical aspects of the exposition are not merely sugar-coating to enable the reader to swallow the material as easily as possible, but are presented for their own inherent interest. Indeed this is not intended to be an "easy" book... by presenting neither an encyclopedic account of all details nor a survey treatment, by concentrating instead on the meaning and the power of some basic ideas... one can hope to leave the student with a view of physical science as an interconnected achievement; one can engage his attention in those lengthier, more rigorous, penetrating arguments and derivations without which an honest understanding of the processes of science is hardly attainable..."

Quoting from the *Report on General Education in a Free Society* (Harvard University Press, 1945, pp. 220–222),

> From the viewpoint of general education the principal criticism... (is that) most of the time... is devoted to developing a technical vocabulary and technical skills... Comparatively little serious attention is given to the examination of basic concepts, the nature of scientific enterprise, the historical development,... its great literature, or its interrelationships with other areas... What such courses frequently supply are only the bricks of the scientific structure. The student who goes on to more advanced work can build

> something with them. The general student is more likely to be left simply with bricks... The emphasis on historical development...is in no sense to constitute a humanistic garnishing of its factual material. On the contrary, it is introduced to illuminate and vitalize the content with which it is integrated.

I have quoted extensively from the prefaces of Holton's books because I think that their point of view would very well serve the needs of liberal arts and other students taking mathematics courses. The ignoring or denying of that point of view by authors and teachers, in my opinion, will lead to further isolation and alienation of mathematics from liberal arts students.

Medical Education. When we consider extending our teaching beyond facts and techniques, we are confronted with felt time pressure. "There is so much to cover, that there is no time for extraneous considerations," is a common response. Medical education, which is notorious for its demands for the memorization of facts, however, is changing.

A former chairman of a Department of Medicine returned to medical school as a full-time student because he was troubled by the problems that he perceived in medical education. He wrote about his experience in the New England Journal of Medicine. [4]

> Learning is a thinking, problem solving process that requires time. Medical school education today involves too little thinking and problem solving. It consists largely of too much fact in too little time....
>
> Let me be clear and not misunderstood. Facts are essential. Problems cannot be solved without the sequential arrangement of facts. But in medicine, the answer may not be there even after the facts are arranged. Students must learn to handle uncertainty: that too is medicine. Emphasis on facts does not teach this aspect of medicine any more than it teaches problem solving.... It is vital to replace the concept of learning as fact gathering to pass examinations with the concept of education as inquisitiveness, sequential thought, problem solving, and the satisfactions that result.

The Association of Medical Colleges issued a report in 1984 ("Physicians for the Twenty-First Century"). Recommendation 1 states,

> In the general professional education of the physician, medical faculties should emphasize the acquisition and development of skills, values and attitudes by students at least to the same extent that they do their acquisition of knowledge. To do this, medical faculties must limit the amount of factual information that students are expected to memorize.

Harvard Medical School (New Pathways) and McMaster University Medical School are two that have decreased lectures and coverage in their effort to improve medical education.

Humanistic Mathematics. Some mathematicians are searching for new approaches to teaching. In March 1986 The Exxon Education Foundation supported a three day Conference to Examine Mathematics as a Humanistic Discipline at Harvey Mudd College, Claremont, California. A common response of the thirty-six mathematicians who participated was "I was startled to see so many who thought as I do".

One theme that emerged was the need to place the student more centrally in the position of inquirer than is generally the case, while at the same time acknowledging the emotional climate of the activity of learning mathematics. A second theme was the need to reconstruct the curriculum and the discipline of mathematics itself. The reconstruction would relate mathematical discoveries to personal courage, relate discovery to verification, mathematics to science, truth to utility, and in general, to relate mathematics to the culture in which it is embedded. There is a movement to continue and broaden the network of mathematicians who would be mutually supportive of efforts to teach and study mathematics according to the themes of the conference.

Mathematics is open ended. Its concepts, methods and contacts with other disciplines are expanding. Easily understood unsolved problems and unanswered questions abound. Philosophic controversy swirls around it. But most students perceive mathematics as a dead, unchanging, irrelevant subject.

Students at every level could come to appreciate that there are unanswered questions as well as unanswerable questions (and to know the difference) by exploring and inventing questions. Mathematics is more than solving problems or proving things. Finding a new idea or a metaphor to describe an old idea, evaluating the significance of concepts and seeing new connections are also reasonable activities in mathematics. **[2]**

REFERENCES

1. Quoted by J. Bronowski, The Ascent of Man, Little, Brown and Co., Boston, 1973.
2. S. I. Brown and M. Walter, The Art of Problem Posing, The Franklin Institute Press, Philadelphia, 1983.
3. E. A. Burtt, The Metaphysical Foundations of Modern Science, Doubleday & Co., New York, 1954.
4. L. W. Eichna, "Medical School Education 1975–79: A student's perspective," *New England Journal of Medicine*, 303 (1980) 729–34.
5. G. Holton, Introduction to Concepts and Theories in Physical Science (second edition), Addison-Wesley, Reading, Mass., 1973.
6. S. K. Stein, "Gresham's Law: Algorithm Drives Out Thought," talk delivered 1/23/87, San Antonio, at special session Mathematics as a Humanistic Discipline, MAA Annual meeting.
7. O. Toeplitz, The Calculus: A Genetic Approach, University of Chicago Press, 1963.
8. A. N. Whitehead, Science and the Modern World, The Free Press, New York, 1967.

TELEGRAPHIC REVIEWS

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T: Textbook	P: Professional Reading	1–4: Semesters
C: Computer Software	L: Undergraduate Library	**: Special Emphasis
S: Supplementary Reading	13: Grade Level	??: Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatabilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

General, L. *Quantum Implications: Essays in Honour of David Bohm.* Ed: B.J. Hiley, F. David Peat. Routledge & Kegan Paul, 1987, vii + 455 pp, $49.95. [ISBN: 0-7102-0806-5] Thirty essays on topics ranging from "Hidden variables and the implicit order" (by Bohm), to "Meaning and being in contemporary physics," to "The automorphism group of C_4," to "Wholeness and dreaming." Quite a statement on the breadth and depth of Bohm's influence on modern science. Contributors include Penrose, Feynman, and R. Weber. MR

Finite Mathematics, T(18). *Mathematics with Applications for the Management, Life, and Social Sciences, Third Edition.* Howard Anton, *et al.* Harcourt Brace Jovanovich, 1988, xv + 1029 pp, $32. [ISBN: 0-15-555237-6] A good introductory book which presents nice motivation for definitions. Includes standard precalculus material; in the final chapters, the authors take an informal and intuitive approach to develop the fundamental ideas and applications of differential and integral calculus, differential equations, and functions of several variables. Exercise sets include drills as well as many word problems. (*Second Edition*, TR, April 1984.) LW

Education, P*, L*. *Computers and Mathematics: The Use of Computers in Undergraduate Instruction.* Ed: David A. Smith, *et al.* MAA Notes No. 9. MAA, 1988, xi + 147 pp, (P). [ISBN: 0-88385-059-1] Nineteen surveys by pedagogical pioneers who are exploring the use of computing across the entire undergraduate curriculum—from calculus to statistics, from differential equations to abstract algebra. A project of the MAA Committee on Computers in Mathematics Education (CCIME), this is no mere committee report, but a stimulating anthology of the best current exemplars on this important frontier of teaching. LAS

History, S, L. *Studies in the History of Mathematics.* Ed: Esther R. Phillips. Stud. in Math., V. 26. MAA, 1987, 308 pp, $32. [ISBN: 0-88385-128-8] "Collection of [ten] articles on modern and contemporary mathematics ... whose purpose is to provide a sample of recent scholarship in the history of mathematics and to suggest additional sources for further study." Each article analyzes a particular event or development in the history of mathematics. BH

History, P, L.** *Gesammelte Abhandlungen, Collected Papers.* Friedrich Hirzebruch. Springer-Verlag, 1987, $195 set [ISBN: 0-387-18087-7]. *Band I, 1951-1962*, viii + 814 pp; *Band II, 1963-1987*, iv + 818 pp. Virtually all of Hirzebruch's papers, in chronological order, with added commentary (at the end of each volume) by the author keyed to occasional marginal indicators to indicate progress since original publication. LAS

Number Theory. *1017 Problems.* A.S. Moiseenko (10-12 Kimball St., Belleville, NJ), 1987, 527 pp, (P). The problems all consist of specific Diophantine equations and their solutions. A photocopy of the author's handwriting. No text; just equations and solutions. CEC

Linear Algebra, T(14-15: 1, 2), S, L*.** *Linear Algebra and Its Applications, Third Edition.* Gilbert Strang. Harcourt Brace Jovanovich, 1988, xii + 505 pp, $30. [ISBN: 0-15-551005-3] "Linear algebra is a fantastic subject." This new edition of a path-breaking text includes a new section on the Fast-Fourier Transform, many new exercises, and a greatly expanded appendix on computer codes, packages, and experiments. In addition, linear transformations are now integrated throughout the text. An exciting book that "allows pure mathematicians to teach applied mathematics" and, in the process, displays as few other texts do the eye-opening power and beauty of linear algebra. (*First Edition*, TR, June-July 1976, Extended Review, January 1978; *Second Edition*, TR, October 1980.) LAS

Algebra, T(18: 2), S, P. *Structure of Blocks of*

Group Algebras. Gregory Karpilovsky. Mono. & Surv. in Pure & Appl. Math., V. 33. Longman Scientific & Technical (US Distr: Wiley), 1987, xviii + 427 pp, $179.95. [ISBN: 0-470-20747-7] A self-contained, up-to-date account of the ring-theoretic structure of blocks of modular group algebras. Ties together the various developments of the last twenty years, and gives a picture of the current state of the subject. Gives a thorough list of references. No exercises. Expensive! CEC

Real Analysis, T(15: 2), S, L*. *Real Analysis with Point-Set Topology.* Donald L. Stancl, Mildred L. Stancl. Pure & Appl. Math., V. 113. Marcel Dekker, 1987, x + 287 pp, $45. [ISBN: 0-8247-7790-5] Gives a basic introduction to point-set topology and metric spaces. Then, covers continuity, sequences and series, sequences of functions, integration, and differentiation in the more general topological or metric space setting while applying the results to the real line. Well motivated with good selection of problems in each section. Good choice for better students. BH

Complex Analysis, T(15-16: 1). *Elements of Complex Analysis.* William J. Adams (Dept. of Math., Pace U., NY 10038), 1987, ix + 337 pp, $27.50 (P). Rigorous approach to first course in complex analysis. Begins with background material on real numbers, sequences of real numbers and complex numbers, basic topological facts, and continuous functions, then carefully develops theory of complex functions. Perhaps too much emphasis on rigorous details and too little on the "big picture" and on giving examples. May be appropriate for more gifted students. BH

Partial Differential Equations, P. *Numerical Simulation in Oil Recovery.* Ed: Mary F. Wheeler. IMA, V. 11. Springer-Verlag, 1988, xi + 283 pp, $25.95. [ISBN: 0-387-96653-6] Proceedings of a December 1986 symposium at the Institute for Mathematics and Its Applications, University of Minnesota. Seventeen papers address a variety of research fields connected with reservoir modelling, including the modelling of fractures, heterogeneities, viscous fingering, and diffusion-dispersion effects. DFA

Numerical Analysis, P. *Lecture Notes in Mathematics-1287: Approximation Theory, Tampa.* Ed: E.B. Saff. Springer-Verlag, 1987, v + 228 pp, $21.20 (P). [ISBN: 0-387-18500-3] Twelve papers from an international seminar held during 1985-86 at the Institute for Constructive Mathematics at the University of South Florida. RWN

Functional Analysis, P. *Lecture Notes in Mathematics-1282: Positive Polynomials, Convex Integral Polytopes, and a Random Walk Problem.* David E. Handelman. Springer-Verlag, 1987, xi + 136 pp, $13.10 (P). [ISBN: 0-387-18400-7] Concerned with results and interconnections in a number of areas including positive polynomials, a class of special random walk problems on lattices, convex integral polytopes, reflection groups, and commutative algebras. The central problems arise from special actions of tori on C^*-algebras. Techniques include those of functional analysis, ordered rings, commutative algebra, and convex analysis. CEC

Analysis, P, L. *A Handbook of Fourier Theorems.* D.C. Champeney. Cambridge U Pr, 1987, xi + 185 pp, $39.50. [ISBN: 0-521-26503-7] "Intended to assist those scientists, engineers, and applied mathematicians who are already familiar with Fourier theory and its applications in a non-rigorous way, but who wish to find out the exact mathematical conditions under which particular results can be used." Begins with an introduction to ideas and terminology of Lebesque integration and then "gives rigorous statements of the most important theorems in Fourier theory with explanatory comments and examples." No proofs; well written. BH

Algebraic Geometry, P. *Algebraic Groups and Class Fields.* Jean-Pierre Serre. Grad. Texts in Math., V. 117. Springer-Verlag, 1988, ix + 207 pp, $36. [ISBN: 0-387-96648-X] A translation of the 1975 French edition. SG

Differential Geometry, T(16-17: 2), L. *Differential Geometry: Manifolds, Curves, and Surfaces.* Marcel Berger, Bernard Gostiaux. Transl: Silvio Levy. Grad. Texts in Math., V. 115. Springer-Verlag, 1988, xii + 474 pp, $52. [ISBN: 0-387-96626-9] Translation of a French text in differential geometry in two parts: an introduction to manifolds and curves with attention to applications of Stokes' formula, non-trivial examples of new notions, and interplay between geometry and analysis (with exercises); and a terser "travel guide" to local and global properties of surfaces in R^3. Presumes "good calculus background, including multivariable calculus and some knowledge of forms in R^n." RB

Differential Geometry, P. *Stochastic and Integral Geometry.* Ed: R.V. Ambartzumian. D Reidel (US Distr: Kluwer Academic), 1987, 136 pp, $34.50. [ISBN: 90-277-2543-8] Seven papers selected from an October 1985 symposium held in Armenia. A reprint of *Acta Applicandae Mathematicae,* Vol. 9, Nos. 1-2, 1987. LAS

Differential Geometry, P. *Global Analysis on Foliated Spaces.* Calvin C. Moore, Claude Schochet. Math. Sci. Res. Inst. Pub., V. 9. Springer-Verlag, 1988, 337 pp, $34. [ISBN: 0-387-96664-1] A foliated space is an n-dimensional manifold which may be partitioned into an $(n-p)$-dimensional family of p-dimensional submanifolds. Foliations arise in the study of flows and dynamics, in group representations, in automorphic forms, and in groups acting on spaces. The book develops a variety of aspects of analysis and geometry on foliated spaces, and contains an exposition of the Connes index theorem for foliated spaces. AM

Differential Geometry, T(17-18: 2), P, L. *Differential Geometry.* Tanjiro Okubo. Pure & Appl. Math., V. 112. Dekker, 1987, xviii + 788 pp, $125. [ISBN: 0-8247-7700-X] Intended as an introduction and reference for local and global differential geom-

etry. Written for graduate students enrolled in differential geometry courses as well as a more general audience which includes geometers, topologists, and mathematical physicists. Among other topics, the book treats differentiable, complex, and Riemannian manifolds, connections, Lie groups, de Rham cohomology, and characteristic classes. Does contain exercises. AM

Differential Geometry, P. *Differential Geometry: The Interface between Pure and Applied Mathematics.* Ed: Mladen Luksic, Clyde Martin, William Shadwick. Contemp. Math., V. 68. AMS, 1987, ix + 273 pp, $29 (P). [ISBN: 0-8218-5075-X] Proceedings of a conference held in San Antonio, Texas, April 23-25, 1986. The book contains sixteen papers covering, among other topics, numerical analysis, mechanics, and control theory. AM

Geometry, S*(14), P, L. *Groups: A Path to Geometry.* R.P. Burn. Cambridge U Pr, 1987, xii + 242 pp, $16.95 (P). [ISBN: 0-521-34793-9] A paperback edition (with corrections) of the 1985 text (TR, June-July 1987). JNC

Geometry, S(14-16), P, L. *The Geometrical Work of Girard Desargues.* J.V. Field, J.J. Gray. Springer-Verlag, 1987, x + 237 pp, $68. [ISBN: 0-387-96403-7] Primarily devoted to a translation of Desargues' *Rough Draft on Conics,* although the authors also provide details of the Greek legacy upon which this work is built. The three-and-one-half century-old work is rendered eminently readable by authors who have used modern notation, given a glossary of Desargues' "botanical vocabulary," and provided a running commentary through elaborate footnotes. They have also introduced their own illustrations, for Desargues' appear to have been lost. Translations are presented of the lesser works on perspective and the sun dial, and a chapter is devoted to Desargues' work on applied geometry. SS

Geometry, S*(16-17), P, L*. *Lectures in Geometry, Semester V: Lie Groups and Lie Algebras.* M. Postnikov. Transl: Vladimir Shokurov. MIR (US Distr: Imported Pub), 1986, 437 pp, $10.95. [ISBN: 0-8285-3296-6] Twenty-one lectures for an elective university course by a great Russian mathematician, with orally-omitted arguments filled in. Basic notions and examples; "local theory" of Lie groups; extending local to global theory; subgroups quotient groups (G_2, F_4 presented); proof of Ado's theorem. Inexpensive yet valuable: a compact, accessible compilation of the major results. RB

Algebraic Topology, P. *Homotopy Theory.* Ed: E. Rees, J.D.S. Jones. London Math. Soc. Lect. Note Ser., V. 117. Cambridge U Pr, 1987, vii + 247 pp, $29.95 (P). [ISBN: 0-521-33946-4] Proceedings of a symposium at Durham University, August 1985, featuring expository papers on recent advances in higher homotopy theory. Ten papers: Burnside ring conjecture; stable splittings; localization and periodicity; global methods, K-theory and homotopy theory; computational complexity of rational homotopy; unstable homotopy since 1950; and symbolic calculus. RB

Algebraic Topology, P. *Lecture Notes in Mathematics-1286: Algebraic Topology.* Ed: H.R. Miller, D.C. Ravenel. Springer-Verlag, 1987, vii + 341 pp, $30.30 (P). [ISBN: 0-387-18481-3] Proceedings of a University of Washington workshop during 1984-85 school year in which a total of 29 topologists participated at various times in four topic courses and a very active seminar. One set of course notes (by F.R. Cohen, on aspects of classical homotopy theory), and eleven research papers are included (Kervaire invariant, loop spaces of spheres, stability, Adams spectral sequences). RB

Topology, P. *Lecture Notes in Mathematics-1283: Geometric Topology and Shape Theory.* Ed: S. Mardešić, J. Segal. Springer-Verlag, 1987, 261 pp, $25.80 (P). [ISBN: 0-387-18443-0] Proceedings of the Postgraduate School and Conference on the subject held at the University of Zagreb, Yugoslavia, October 1986. Twenty research papers on topics including decomposition theory, cell-like mappings, infinite-dimensional spaces, shape fibrations, fibered shape, LC^n-compacta, generalized manifolds, embedding continua into manifolds, complement theorems in shape theory, shape-theoretic methods in group theory, exact homologies, strong shape theory. RB

Topology, S(16-18), P. *On Knots.* Louis H. Kauffman. Annals of Math. Stud., No. 115. Princeton U Pr, 1987, xv + 480 pp, $50; $18.95 (P). [ISBN: 0-691-08434-3; 0-691-08435-1] Expanded version of notes from a seminar held at the University of Zaragoza, Spain, Winter 1984. Elementary concepts of diagram moves, linking numbers; knot cobordism and the Arf invariant; geometric knot theory with covering spaces and branched covers; topology of singularities and Brieskorn varieties; collection of miscellaneous topics compiled throughout the course. RB

Control Theory, T(18), P. *Game-Theoretical Control Problems.* N.N. Krasnovskiĭ, A.I. Subbotin. Transl: Samuel Kotz. Springer-Verlag, 1987, xi + 517 pp, $125. [ISBN: 0-387-96389-8] Translated, substantially-revised version of the authors' Russian monograph *Positional Differential Games* (1974). "[A] formalization of differential games is presented, existence theorems for equilibrium are proved, ..., attention is devoted to the construction of optimal strategies, ..., stability of proposed solutions ...." RB

Probability, P. *Multivariate Empirical Processes.* J.H.J. Einmahl. CWI Tract, V. 32. Math Centrum, 1987, iv + 99 pp, Dfl. 14.10 (P). [ISBN: 90-6196-312-5] A research tract dealing with a classical approach to weighted empirical processes based on independent, identically distributed random variables. Includes a discussion of weak convergence and strong limit theorems. Extensive bibliography. TAV

Elementary Statistics, T(7-12), S**.** *Quantitative Literacy Series.* Dale Seymour Pub (POB 10888, Palo Alto, CA 94303). *Exploring Data.* James M. Landwehr, Ann E. Watkins. 1986, 161 pp,

$9.95 (P), [ISBN: 0-86651-321-3]; *Exploring Probability.* Claire M. Newman, Thomas E. Obremski, Richard L. Scheaffer. 1987, 63 pp, $7.95 (P), [ISBN: 0-86651-333-7]; *The Art and Techniques of Simulation.* Mrudulla Gnanadesikan, Richard L. Scheaffer, Jim Swift. 1987, 53 pp, $7.95 (P), [ISBN: 0-86651-336-1]; *Exploring Surveys and Information from Samples.* James M. Landwehr, Jim Swift, Ann E. Watkins. 1987, 96 pp, $8.95 (P). [ISBN: 0-86651-339-6] Four volumes of innovative workbooks developed by ASA and NCTM to introduce statistical methods into mathematics courses in grades 7-12. *Exploring Data* covers stem-and-leaf plots, box plots, scatterplots, smoothing, and other visual representations of data; examples fit the world view of young teenagers. *Exploring Probability* uses experiments (e.g., tree diagrams, guessing numbers, throwing darts to estimate areas) to develop intuition about chance leading to simple examples of complementary and compound events. *Simulation* introduces computer simulation of coin tosses and random numbers, then follows with mini-projects involving a wide variety of waiting-time applications. *Exploring Surveys,* intended for high school students, explores sampling, confidence intervals, capture-recapture methods, and other techniques in the context of extensive case studies and student projects. Steep discounts available for class purchases; teacher's edition, software, and videotape supplement also available. LAS

Computational Statistics, T(14-17: 1), S. *A Guide to SPSS/PC+.* Neil Frude. Springer-Verlag, 1987, xvi + 279 pp, $27 (P). [ISBN: 0-387-91312-2] Tutorial introduction to the Statistical Package for the Social Sciences (SPSS) as adapted for use on IBM PC/XT or IBM PC/AT or equivalent personal computers. Focuses on data collected in a questionnaire survey, following the process through to the production of a final report. Prepares the reader to cope with the more technical *SPSS/PC+ Manual.* RSK

Statistics, P. *Cancer Modeling.* Ed: James R. Thompson, Barry W. Brown. Stat.: Textbooks & Mono., V. 83. Marcel Dekker, 1987, xii + 422 pp, $89.75. [ISBN: 0-8247-7773-5] A series of ten articles in which fifteen leading experts elucidate methods to describe the genises, progression, results, and the treatment of cancer. Methodologies that range from purely descriptive to highly sophisticated are demonstrated. Includes suggestions for the extension of current methods. CEC

Statistics, P. *Coordinate-Free Multivariable Statistics: An Illustrated Progression from Halmos to Gauss and Bayes.* Mervyn Stone. Stat. Sci. Ser., V. 2. Clarendon Pr, 1987, xiv + 120 pp, $29.95. [ISBN: 0-19-852210-X] Uses a coordinate-free geometric approach to provide a theoretical framework for linear and affine multivariate statistical methods. Relies heavily on linear algebra, particularly dual vector spaces, aided by a novel way to represent these concepts pictorially. RSK

Statistics, T(16-17: 1, 2), P, L. *Linear Models for Unbalanced Data.* Shayle R. Searle. Prob. & Math. Stat. Wiley, 1987, xxiv + 536 pp, $49.95. [ISBN: 0-471-84096-3] First half presents linear model methodology without using matrix algebra for data having unequal numbers of observations in the subclasses. Second half uses matrices to extend these concepts. Deals only with the fixed effects model, except for the last chapter which gives a summary review of the mixed model. Includes comments on computer output from five of the major statistical packages. RSK

Statistics, P. *Topics in Multivariate Approximation.* Ed: C.K. Chui, L.L. Schumaker, F.I. Utreras. Academic Pr, 1987, x + 335 pp, $39.95. [ISBN: 0-12-174585-6] Twenty-one survey papers delivered at a December 1986 workshop at the University of Chile in Santiago. Topics include multivariate splines, fitting of scattered data, tensor approximation methods, multivariate polynomial approximation, numerical grid generation, finite element methods, constrained interpolation and smoothing. Also includes a bibliography of over 1100 entries by L.L. Schumaker and R. Franke. DFA

Statistics, T*(17: 1, 2). *Applied Multivariate Statistical Analysis, Second Edition.* Richard A. Johnson, Dean W. Wichern. Ser. in Stat. Prentice-Hall, 1988, xvi + 607 pp, $50.33. [ISBN: 0-13-041146-9] Revision of the authors' 1982 *First Edition* (TR, December 1982). Main changes include the addition of a chapter on canonical correlation analysis, and expansion of the chapter on factor analysis to include some recent material on computer-aided analysis of structural equations. RSK

Languages, P. *Lecture Notes in Computer Science-285: Semantics of Digital Circuits.* Carlos Delgado Kloos. Springer-Verlag, 1987, ix + 124 pp, $16.40 (P). [ISBN: 0-387-18540-2] "The theory of formal semantics for programming languages was developed for precisely defining the function of computer programs .... It has now become evident that a similar treatment is necessary for digital circuits." A language for describing concurrent systems is defined and applied, with different domains, to various levels of VLSI design. RB

Languages, P. *Lecture Notes in Computer Science-283: Category Theory and Computer Science.* Ed: D.H. Pitt, A. Poigné, D.E. Rydeheard. Springer-Verlag, 1987, v + 300 pp, $27.30 (P). [ISBN: 0-387-18508-9] Proceedings of a conference held at Edinburgh, United Kingdom, September 1987, describing links being established between category theory and computer programming. Sixteen papers; in some, category theory relates different approaches to a topic (e.g., axiomatic and denotational semantics); others introduce category theoretic models (e.g., of logic programming, of distributed computing). RB

Languages, P. *Lecture Notes in Computer Science-288: MetaSoft Primer.* Andrzej Blikle. Springer-Verlag, 1987, xiii + 140 pp, $19.10 (P). [ISBN: 0-387-18657-3] In denotational semantics, a function which associates a value with each syntactic construct in a programming language determines the meaning of a program. This book covers mathematical frame-

work for the author's simplified denotational semantics theory, discusses application to an example language (Pascal), and reports on development of an automated system MetaSoft for implementing applications. RB

Computer Systems, P. *TRON Project 1987: Open-Architecture Computer Systems.* Ed: Ken Sakamura. Springer-Verlag, 1987, xi + 312 pp, $59.50. [ISBN: 0-387-70027-7] The editors proposed 'The Realtime Operating System Nucleus' concept five years ago. These are proceedings of the third symposium of the international multicorporate association subsequently formed to carry out the project. Twenty-three papers report on phases of this effort "to bring the dream of computerized society ... into reality:" industrial applications (extended to the home); workstations; servers, gateways; networks; the unifying VLSI chip. RB

Computer Systems, P. *The Amoeba Distributed Operating System: Selected Papers 1984-1987.* Ed: Sape J. Mullender. CWI Tract, V. 41. Math Centrum, 1987, vi + 309 pp, Dfl. 46.70 (P). [ISBN: 90-6196-325-7] A collection of previously-published articles on Amoeba, which is a joint effort of the Center for Mathematics and Computer Science and Vrije University, both in Amsterdam. Nineteen papers by Tanenbaum, van Renesse, Mullender, and others comment on aspects of protection, protocols, reliability, file systems, wide area networking, applications, experience. RB

Computer Systems. *PasRo: Pascal and C for Robots, Second Extended Revision.* C. Blume, W. Jakob, J. Favaro. Springer-Verlag, 1987, ix + 239 pp, $35 (P). [ISBN: 0-387-18093-1] Aspects of this robot programming system are new since the *First Edition* (TR, April 1986). Includes the PasRo description of that edition (based on Pascal), revised and extended, and a new PasRo description based on the programming language C. DFA

Computer Systems, T(13-17: 1), S, P, L. *Computer Communications, Second Edition.* Robert Cole. Springer-Verlag, 1987, viii + 173 pp, $22 (P). [ISBN: 0-387-91306-8] How do computers communicate? Beginning with transmission in wires and ending with computer network protocols and end-to-end services, this brief text surveys its topic with an easily followed mixture of theory, technique, and example. Intended for undergraduates in computer science, advanced computer studies students; suitable for computer professionals. RB

Computer Systems, L, P. *Computing and Change on Campus.* Ed: Sara Kiesler, Lee Sproull. Cambridge U Pr, 1987, xiii + 255 pp. [ISBN: 0-521-34431-X] A series of research papers profiling the impact of computing at Carnegie-Mellon University during the period 1982-1986, examining attitudes, behavior, effect on employees (secretaries, librarians, managers), and on students. Little is said about the impact on curriculum or on learning. Concluding chapter summarizes present CML policies based on this extraordinary period of rapid computer growth. LAS

Computer Graphics, P, L. *Color and the Computer.* Ed: H. John Durrett. Academic Pr, 1987, x + 299 pp, $59.50. [ISBN: 0-12-225210-1] Motivated by experience with inconsistent, incompatable, distracting use of color, the editor assembled fundamental information on color, color computer displays and human color perception, as a step toward more effective use. In addition, eleven application areas (including medical images, cockpit instrumentation, map making, business graphics) are explored. RB

Computer Graphics, T(15-18: 1, 2), S, P, L. *An Introduction to Solid Modeling.* Martti Mäntylä. Princip. of Comput. Sci. Ser., V. 13. Computer Science Pr, 1988, xiv + 401 pp, $42.95. [ISBN: 0-88175-108-1] Solid modeling refers to representing and processing geometric information on three-dimensional solid objects. Part one presents background, problems, models employed in solid modeling. Case studies of algorithms are given. The second part is the *The Geometric Workbench,* a simple solid modeling system based on the boundary modeling approach. Sample programs in C. Problem sections. Appendices on homogeneous coordinates and point-set topology. Bibliography, list of programs, index. RJA

Artificial Intelligence, S(16-17), P, L. *Artificial Intelligence and Tutoring Systems: Computational and Cognitive Approaches to the Communication of Knowledge.* Etienne Wenger. Morgan Kaufmann, 1987, xxiii + 486 pp, $32.95. [ISBN: 0-934613-26-5] Survey and conceptual analysis of intelligent tutoring systems (ITS). Focus on the perspective of communication of knowledge (warranted belief) in a shift from CAI (where emphasis is on programming of expert's decisions) to the programming of knowledge necessary for ITS. Underlying models of the student, of knowledge representation, and communication are explored as a basis for a cognitively-oriented software engineering. RM

Artificial Intelligence, P. *Expert Judgment and Expert Systems.* Ed: Jeryl L. Mumpower, *et al.* NATO ASI Ser. F, V. 35. Springer-Verlag, 1987, viii + 361 pp, $59.50. [ISBN: 0-387-17986-0] Proceedings of a NATO advanced research workshop, August 1986. Researchers in psychology, decision analysis, and artificial intelligence gathered to compare the ways their disciplines study expert judgment, to evaluate relative strengths and weaknesses, and to propose profitable linkages among the approaches. Four keynote addresses, fifteen invited presentations. RB

Artificial Intelligence, P. *Machine Intelligence and Knowledge Engineering for Robotic Applications.* Ed: Andrew K.C. Wong, Alan Pugh. NATO ASI Ser. F, V. 33. Springer-Verlag, 1987, xiv + 486 pp, $79. [ISBN: 0-387-17844-9] Proceedings of the NATO Advanced Research Workshop held at Maratea, Italy, May 12-16, 1986. A collection of papers which review recent advances in machine intelligence and knowledge engineering for robotic applications. Covers robot vision, knowledge representation and image understanding, robot control and inference systems, task planning and expert systems.

and integrated software/hardware systems. Presentations include system implementation and industrial applications issues. PS

Computer Science, P. *Lecture Notes in Computer Science-282: Visualization in Programming.* Ed: P. Gorny, M.J. Tauber. Springer-Verlag, 1987, 210 pp, $21.80 (P). [ISBN: 0-387-18507-0] Proceedings of a Fifth Workshop on Informatics and Psychology held in Schärding, Austria, May 1986. Thirteen papers by European and American authors, including a survey of visual languages (those programmed in visual images *and* those for producing visual images), several psychology research papers (e.g., experiment comparing BOXER visual LOGO with ordinary LOGO; cognitive analysis of visual interfaces), interface design issues. RB

Computer Science, S(18), P. *Lecture Notes in Computer Science-281: Trends, Techniques, and Problems in Theoretical Computer Science.* Ed: Alica Kelemenová, Jozef Kelemen. Springer-Verlag, 1987, vi + 213 pp, $21.80 (P). [ISBN: 0-387-18535-6] Selected contributions to the Fourth International Meeting of Young Computer Scientists, Smolenice, Czechoslovakia, October 1986. VLSI algorithms, problems in formal language theory, various aspects of the theory of formal grammars, biologically-motivated structures, topics in artificial intelligence. DFA

Applications, P. *Maximum-Entropy and Bayesian Spectral Analysis and Estimation Problems.* Ed: C. Ray Smith, Gary J. Erickson. Fundament. Theor. of Physics. D Reidel (US Distr: Kluwer Academic), 1987, ix + 322 pp, $59. [ISBN: 90-277-2579-9] Proceedings of a workshop held at the University of Wyoming in August 1983; includes papers applying Bayesian methods in quite diverse areas (e.g., signal processing, theory of the brain, tomography, databases). LAS

Applications (Electrical Engineering), T(16-17: 1), L. *The Theory of Electromagnetic Flow-Measurement.* J.A. Shercliff. Cambridge U Pr, 1987, xiii + 146 pp, $14.95 (P). [ISBN: 0-521-33554-X] Reprint without changes of the text first published in 1962. The theory of electromagnetic flow measurement concerns the determination of fluid flow rates through measurement of the electromagnetic flow induced by a fluid moving through a magnetic field. AM

Applications (Management), T(13: 1-3). *Mathematics with Applications in Management and Economics, Sixth Edition.* Earl K. Bowen, Gordon D. Prichett, John C. Saber. Irwin, 1987, xxiii + 993 pp, $37.95. [ISBN: 0-256-03140-1] Accessible to students with only one year of secondary school algebra, this text includes the prerequisite mathematics required for its coverage of linear programming, mathematics of finance, simple optimization, and elementary probability and statistics. Computer programs in Basic and Pascal and the use of computer packages such as LINDO and Minitab are included. (*Third Edition,* TR, May 1973; *Fourth Edition,* TR, December 1976; *Fifth Edition,* TR, October 1980.) JNC

Applications (Physics), P. *A Commentary on Thermodynamics.* William Alan Day. Tracts in Nat. Philo., V. 32. Springer-Verlag, 1988, ix + 96 pp, $48. [ISBN: 0-387-96615-3] The objective here is to set some topics from thermoelasticity in the context of field theories which incorporate the effects of heat conduction and inertia. The topics include linear and nonlinear thermoelasticity. MU

Applications (Physics), T(18: 1, 2), P. *Path Integral Methods in Quantum Field Theory.* R.J. Rivers. Mono. on Math. Physics. Cambridge U Pr, 1987, xi + 339 pp, $69.50. [ISBN: 0-521-25979-7] A concise graduate-level introduction to analytic functional methods in quantum field theory. The unusually well-written, clear, and engaging text is heavily laced with elegant mathematics. MU

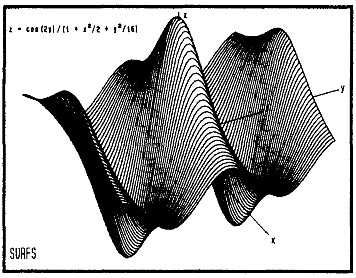
Applications (Physics), S(16-18), P, L. *Spinors and Space-Time, Volume 1: Two-Spinor Calculus and Relativistic Fields.* Roger Penrose, Wolfgang Rindler. Mono. on Math. Physics. Cambridge U Pr, 1987, x + 458 pp, $29.95 (P). [ISBN: 0-521-33707-0] The two-spinor calculus replaces the usual tensor calculus in this formal analysis of four-dimensional space-time geometry. The text begins at a relatively elementary level and progresses carefully through such topics as Einstein-Maxwell equations and Yang-Mills fields. MU

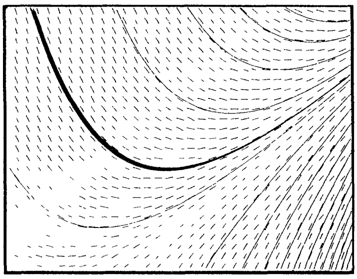
Applications (Physics), T*(16-17: 1), S, P, L*. *Introduction to Modern Statistical Mechanics.* David Chandler. Oxford U Pr, 1987, xiv + 274 pp, $21.95 (P); $39.95. [ISBN: 0-19-504277-8] An excellent introduction emphasizing major modern topics such as Monte Carlo sampling, renormalization groups, and the fluctuation-dissipation theorem. Exercise 2.18 is a must: "Put some water in a bowl. Sprinkle pepper on its surface. Touch a bar of soap to the surface in the middle of the bowl. What happened? Touch the surface with the soap again, and again. What happened and why? (Hint: Consider the formula for the Gibbs adsorption isotherm.)" BC

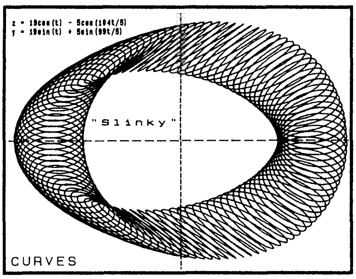
Reviewers

RJA: Richard J. Allen, St. Olaf; DFA: David F. Appleyard, Carleton; RB: Richard Brown, Carleton; LB-E: Lisa Brown-Ewing, Macalester; JNC: Judith N. Cederberg, St. Olaf; LC: Laura Chihara, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; RSF: Robert S. Fisk, St. Olaf; SG: Steven Galovich, Carleton; GG: George Gilbert, St. Olaf; BH: Bruce Hanson, St. Olaf; RH: Rhonda Hatcher, St. Olaf; RK: Roger Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; AM: Alan Magnuson, St. Olaf; SM: Steve McKelvey, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; MR: Matthew Richey, St. Olaf; AWR: A. Wayne Roberts, Macalester; GMS: G. Michael Schneider, Macalester; PS: Paul Schoening, St. Olaf; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MS: Myriam Steinback, Macalester; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; LW: Lynne Walling, St. Olaf; FLW: Frank L. Wolf, Carleton; MZ: Michael Zielinski, St. Olaf.

THE BRIDGE SOFTWARE MATH UTILITIES







SURFS

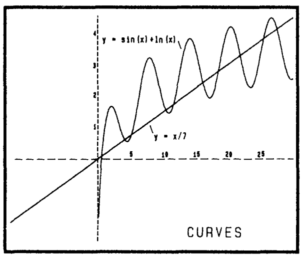
SURFS makes plotting functions z = f(x,y) in three dimensions easy and effective. You can opt for automatic scaling and positioning, but all aspects of the plot, including rotations through any angles, stretchings and translations are modifiable from simple menus. Hidden-line removal and correct axis placement are built in; cross hatching is another option. You can also set the number and the fineness of contours, and change the domain of the function.

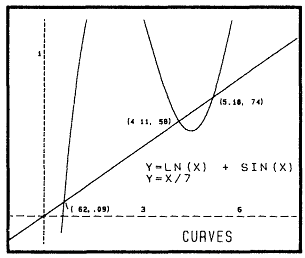
DIFFS

DIFFS plots the direction fields and solution curves to ordinary first order differential equations y′ = f(x,y).

CURVES

This is a very powerful yet easy-to-use curve sketcher. You can enter equations in Rectangular, Polar, or Parametric form and CURVES will plot them. Scaling is possible for special effects. CURVES also provides a unique movable "magnification window" permitting repeated "zooming in" on any section of the plot. You can use this to find the solution of simultaneous equations to a high degree of accuracy.





FEATURES:
- High Resolution output to a dot matrix printer: 1000 x 1000 dot images are **Near Plotter Quality**
- Easy Keyboard input of function
- Full 8087 support
- Easy to understand documentation, with examples and illustrations. No programming necessary.

REQUIRES:
- **IBM PC, XT, AT** or compatibles, with/without **8087** chip; **320K** RAM; **PC-DOS, MS-DOS,** versions 2.0 or higher.
- **PRINTER:** Standard version supports Epson, IBM or compatible dot-matrix printers. Special versions available for Toshiba, Gemini, NEC and C-Itoh.
- **Plotter:** HP plotter version (SURFS, CURVES) for slight extra charge.

"They have been a delight to use."
"I just can't imagine being without it."
- Anthony Barcellos, **PC Magazine**

CALL FOR OUR EDUCATOR'S SPECIAL All 3 Programs Only... **$99** $125 Value

BRIDGE SOFTWARE EDUCATIONAL SCIENTIFIC WARE

P.O.118, New Town Branch, Boston, MA 02258 (617) 527-1585

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 95, Number 8 October 1988

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Dept. of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Shorter articles about the teaching of mathematics should be sent to either MELVIN HENRIKSEN (linear algebra, algebra, differential equations), Department of Mathematics, Harvey Mudd College, Claremont, CA 91711, or STAN WAGON (calculus, analysis, number theory, discrete mathematics, statistics), Department of Mathematics, Smith College, Northampton, MA 01063. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to ANITA E. SOLOW, Dept. of Mathematics, Grinnell College, Grinnell, IA 50112; in geometry or topology to DENNIS DETURCK, Dept. of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to DAVID J. HALLENBECK, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Proposed problems (three copies) and solutions (two copies), both elementary and advanced to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, University of Calgary, Alberta, Canada T2N 1N4.

Reprint Permission: A. B. WILLCOX, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Back Issues: P. and H. BLISS Co., Middletown, CT 06457.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $26 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $29. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $90 per year.

Second class postage paid at Washington, DC, and additional mailing offices.

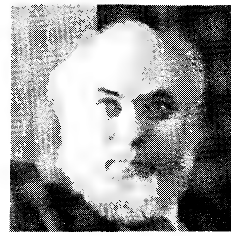
Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Assocation of America, 1529 Eighteenth Street, N.W., Washington, DC 20077-9564.

PRINTED IN THE UNITED STATES OF AMERICA

Special Functions and the Bieberbach Conjecture

NICHOLAS D. KAZARINOFF, SUNY at Buffalo

NICHOLAS D. KAZARINOFF: I received my Ph.D. for a thesis on special functions under the direction of the late R. E. Langer in 1954 at the University of Wisconsin. I have taught at Purdue University, the University of Michigan and the State University of New York at Buffalo. My current interests include fluid dynamics, dynamical systems, iteration of real and complex valued maps, and the geometry of theta series.



In [6] J. Korevaar showed how the inequality

$$ {}_3F_2\left(-n^* + k + 1, k + \tfrac{1}{2}, n^* + k + 1; k + \tfrac{3}{2}, 2k + 1; e^{-t}\right) > 0, \qquad (t > 0,\ n^* \geqslant 2,\ k = 1, 2, \ldots, n^* - 1), \tag{1} $$

which is implied by the result

$$ F(x, \alpha, n) \equiv {}_3F_2\left(-n, n + \alpha + 2, \tfrac{1}{2}(\alpha + 1); \alpha + 1; \tfrac{1}{2}(\alpha + 3); \tfrac{1}{2}(1 - x)\right) > 0 \qquad (-1 < x \leqslant 1,\ n = 0, 1, 2, \ldots, \alpha > -2) \tag{2} $$

of R. Askey and G. Gasper [1], is used by L. de Branges in his proof [4] that the celebrated Bieberbach conjecture is correct. My goal in this article is to give a largely self-contained exposition of Askey and Gasper's proof of (2), accessible to all mathematicians and graduate students who are interested in learning the crucial role that special functions played in the proof of Bieberbach's conjecture. In their paper [1] Askey and Gasper actually prove much more than (2). Here I prove less; it suffices to prove (2) for $\alpha > -1$.

The proof of (2) I give in section 4 below for $\alpha > -1$ reduces to three essential ingredients. The first one of these is the identity

$$ {}_3F_2\left(2a, 2b, a + b; 2a + 2b, a + b + \tfrac{1}{2}; x\right) = \left[{}_2F_1\left(a, b; a + b + \tfrac{1}{2}; x\right)\right]^2, \tag{3} $$

proved in 1828 by Th. Clausen in Altona (now a suburb of Hamburg, F.R.G.) in a short paper in Crelle's Journal [2]. The second is an identity for the Gegenbauer (or ultraspherical) polynomials found in L. K. Hua's book [5, p. 141]; namely, for $\nu > \lambda > 0$

$$ C_n^{\nu}(x) \equiv \sum_{k=0}^{[n/2]} c_k C_{n-2k}^{\lambda}(x), \tag{4} $$

where $[n/2]$ is the greatest integer in $n/2$ and

$$ c_k(\nu, \lambda) = \frac{(n - 2k + \lambda)\Gamma(\lambda)(\nu - \lambda)_k \Gamma(n + \nu - k)}{k!\Gamma(\nu)\Gamma(n + \lambda - k + 1)}. \tag{5} $$

This identity likely can be traced back to Gegenbauer. His polynomials are

conveniently defined by the generating function relationship

$$(1-2xt+t^2)^{-\nu} = \sum_0^\infty C_n^\nu(x)t^n. \tag{6}$$

Since, by ([7, p. 141], [8, p. 220])

$$C_n^\nu(x) \equiv [(2\nu)_n/n!]\,{}_2F_1(-n, n+2\nu; \nu+\tfrac{1}{2}; \tfrac{1}{2}(1-x)), \tag{7}$$

(4)–(5) can be rewritten as

$$\begin{aligned}&{}_2F_1(-n, n+2\nu; \nu+\tfrac{1}{2}; \tfrac{1}{2}(1-x))\\ &\quad= \sum_{j=0}^{[n/2]} d_j(\nu,\lambda)\,{}_2F_1(2j-n, n-2j+2\lambda; \lambda+\tfrac{1}{2}; \tfrac{1}{2}(1-x)),\end{aligned} \tag{8}$$

where

$$d_j(\nu,\lambda) = [n!(2\lambda)_{n-2j}/(n-2j)!(2\nu)_n]c_j. \tag{9}$$

The third ingredient, implicit in Askey and Gasper's proof, is a transformation $T_{e,c}$ that acts as follows:

$${}_3F_2(a,b,c;d,e;t) = T_{e,c}[{}_2F_1(a,b;d;t)] \qquad (e > c > 0). \tag{10}$$

With $\alpha = 2k$, $n = n^* - k - 1$, and $x = 1 - 2e^{-t}$, (2) becomes (1). The conditions on t, k and n^* for the inequality (1) to hold are clearly fulfilled if $-1 < x \leqslant 1$, $\alpha > -1$, and n is a nonnegative integer. I shall define $T_{e,c}$ and prove (2) in section (4) after I prove Clausen's identity (3) in section (2) and Gegenbauer's identity (4)–(5) in section 3. Upon completing this Introduction, I invite the reader to skip sections (2) and (3) at first reading and go directly to section (4) to see how the ingredients described above are simply combined to establish the inequality (2).

For those unfamiliar with the special functions used above and later I close this Introduction with a presentation of the leading characters. First, the generalized factorial $(a)_n = a(a+1)\cdots(a+n-1)$ for $n = 1, 2, \ldots,$ and $(a)_0 = 1$. The Gaussian hypergeometric function ${}_2F_1(a, b; d; x)$ appearing above is the solution of Gauss's hypergeometric differential equation

$$x(1-x)y'' + [d-(a+b+1)]xy' - aby = 0 \tag{11}$$

with $y(0) = 1$ and $y'(0) = ab/d$. Similarly, the generalized hypergeometric function ${}_3F_2(a', b', c'; d', e'; x)$ is the solution of

$$\begin{aligned}&x^2(1-x)z''' + [(3+a'+b'+d')x^2 - (1+c'+e')x]z''\\ &\quad+[c'e' - (1+a'+b'+d'+a'b'+a'd'+b'd')x]z' - a'b'c'z = 0,\end{aligned} \tag{12}$$

with $y(0) = 1$, $y'(0) = a'b'c'/d'e'$, and $y''(0) = (a'+1)(b'+1)(c'+1)/(d'+1)(e'+1)$. These functions have the power series representations implied by the definition

$${}_pF_q(a_1,\ldots,a_p;b_1,\ldots,b_q;x) = \sum_{n=0}^\infty \prod_{i=1}^p (a_i)_n x^n \Big/ \left[n!\prod_{i=1}^q (b_i)_n\right]. \tag{13}$$

I show at the end of the next section that appropriate specializations of the ${}_pF_q$'s in (13) satisfy the equations (11) and (12).

It follows from (6) (see also [7, p. 139], [8, p. 219]) that

$$\begin{aligned}(1-2xt+t^2)^{-\nu} &= \sum_{n=0}^{\infty}(\nu)_n(2xt-t^2)^n/n!\\ &= \sum_{n=0}^{\infty}\sum_{k=0}^{n}(\nu)_n(2x)^{n-k}t^{n+k}/[n!(n-k)!],\\ &= \sum_{n=0}^{\infty}t^n\sum_{k=0}^{[n/2]}(-1)^k(\nu)_{n-k}(2x)^{n-2k}/[k!(n-2k)!].\end{aligned}$$

Thus,

$$C_n^{\nu}(x) = \sum_{k=0}^{[n/2]}(-1)^k(\nu)_{n-k}(2x)^{n-2k}/[k!(n-2k)!]. \tag{14}$$

The Gegenbauer polynomials are orthogonal over $(-1,1)$ with weight function $(1-x^2)^{\nu-1/2}$, so that, in particular, any positive integral power of x may be expanded in a finite sum [5, p. 141; 7, p. 227; 8, p. 281–283]:

$$(2x)^p/p! = \sum_{k=0}^{[p/2]}(-1)^k(\lambda+p-2k)C_{p-2k}^{\lambda}(x)/[k!(\lambda)_{p-k+1}]. \tag{15}$$

I outline a proof of (15) at the beginning of section 3. Actually orthogonality is not what makes an expansion like (15) hold, but rather it is that the $C_n^{\nu}(x)$ $(n = (0,1,\dots)$ form a *simple* set. A set of polynomials is simple if there is exactly one polynomial of each nonnegative degree in the set. Hence any polynomial $P_N(x)$ of degree N can be written as a linear combination of the polynomials in the set $\{p_n(x)\}$: just subtract the multiple c of p_N that cancels the leading term in P_N. Then do the same with $P_N - cp_N$, etc.

2. Clausen's proof of (3). Clausen's identity (3) lies at the root of R. Askey and G. Gasper's proof that the members of a certain family of ${}_3F_2$'s are positive. The proof of (3) given by Clausen in [2] is straightforward. He also proved in [2], [3] that (3) is the only case where a $[{}_2F_1]^2$ of argument x is a constant multiple of a ${}_3F_2$ with the same argument. Thus, one cannot expect to find a simpler proof of (2) than the one given in section 4. Clausen's proof of (3) is as follows. (I omit some tedious elementary algebra.)

Clausen's basic idea is: Suppose that u and v satisfy a linear second order differential equation $L[y] \equiv py'' + qy' + ry = 0$ and answer the question "what third order equation is satisfied by u^2, uv, and v^2?" by determining A, B, P, Q, R, and S so that with $z = u^2$ $(Au + Bu')L[u] + 2u(L[u])' = M[z] \equiv Pz''' + Qz'' + Rz' + Sz$ is an identity. For the case considered by Clausen this procedure is changed a little, because the third order equation to be satisfied by u^2, uv, and v^2 is to have the particular form of an equation satisfied by a ${}_3F_2$. Nevertheless, it succeeds.

To carry out Clausen's idea, multiply both sides of the differential equation (11) for $y = {}_2F_1(a, b; c; x)$ by x to obtain

$$xL[y] \equiv x^2(1-x)y'' + [cx - (a+b+1)x^2]y' - abxy = 0. \tag{16}$$

Then differentiate (16) to obtain

$$(xL[y])' \equiv x^2(1-x)y''' + [(c+2)x - (a+b+4)x^2]y'' + [c - (2a+2b+ab+2)x]y' - aby = 0. \tag{17}$$

On the other hand, $z = {}_3F_2(a', b', d'; c', e'; x)$ satisfies the differential equation (12). Call its left-hand side $M[z]$. But if $z = y^2$, then

$$z' = 2yy', \quad z'' = 2yy'' + 2(y')^2, \quad \text{and} \quad z''' = 2yy''' + 6y'y''. \tag{18}$$

Replace z, z', z'', and z''' in (12) by their equivalents in (18). The left-hand side of the result is $M[y^2]$. Then consider

$$(2Ay + Bxy')L[y] + 2y(xL[y])' \equiv M[y^2]. \tag{19}$$

If $y = {}_2F_1(a, b; c; x)$, the left-hand side is zero. Then $M[y^2] = 0$, provided (19) is an identity; and we obtain (3). To force (19) to be an identity, think of both sides of (19) as being polynomials in y, y', y'', y'''. These polynomials have no common factors. Thus (19) is an identity if the coefficients of corresponding monomials in powers of y, y', y'', y''' are equal. Call the resulting system of equations (S). To simplify the algebra used for computing a', b', c', ... Clausen introduced auxiliary variables $r = a' + b' + c'$, $s = a'b' + a'c' + b'c'$, $u = a'b'c'$, $v = d' + e'$, and $w = d'e'$. In terms of these variables (S) becomes

$$3 + r = A + b + 4, \quad 1 + v = A + c + 2,$$

$$1 + r + s = A(a+b+1) + \tfrac{1}{2}Bab + 2(a+b+1) + ab,$$

$$w = c(A+1), \quad u = 2ab(A+1), \quad 6 = B, \quad 2(3+r) = B(a+b+1),$$

$$2(1+v) = Bc.$$

It easily follows that

$$B = 6, \quad A = 2(a+b) - 1, \quad r = 3(a+b), \quad c = a + b + \tfrac{1}{2}, \quad v = 3c - 1,$$

$$w = (2c-1)c, \quad u = 4(a+b)ab, \quad \text{and} \quad s = 2(a+b)^2 + 4ab. \tag{20}$$

One now concludes from the definitions of $r, \ldots, w$ and (19) that a', b' and c' are the zeros $2a$, $2b$, and $a + b$ of the cubic $\gamma^3 - r\gamma^2 + s\gamma - u$. Further, d' and e' are the zeros c and $2c - 1$ of the quadratic $\gamma^2 - v\gamma + w$. Thus (3) is established.

In terms of the operator $\theta = xd/dx$ the differential equation (11) becomes $[\theta(\theta + c - 1) - x(\theta + a)(\theta + b)]w = 0$. This suggests that the function $w = {}_pF_q$ in (13) satisfies

$$\left[\theta \prod_{i=1}^{q} (\theta + b_i - 1) - x \prod_{i=1}^{p} (\theta + a_i)\right] w = 0. \tag{21}$$

To see this is the case observe that since $\theta x^k = kx^k$,

$$\theta \prod_{i=1}^{q} (\theta + b_i - 1) w = \sum_{k=1}^{\infty} \left[k \prod_{i=1}^{q} (k + b_i - 1) \prod_{i=1}^{p} (a_i)_k x^k\right] \Big/ \left[k! \prod_{i=1}^{q} (b_i)_k\right]$$

$$= \sum_{k=1}^{\infty} \left[\prod_{i=1}^{p} (a_i)_k x^k\right] \Big/ \left[(k-1)! \prod_{i=1}^{q} (b_i)_{k-1}\right].$$

Replace k by $k+1$ to get

$$\begin{aligned}\theta \prod_{i=1}^{q}(\theta+b_i-1)w &= \sum_{k=0}^{\infty}\left[\prod_{i=1}^{p}(a_i)_{k+1}x^{k+1}\right]\Big/\left[k!\prod_{i=1}^{q}[(b_i)_k\right]\\ &= \sum_{k=0}^{\infty}\left[\prod_{i=1}^{p}(a_i+k)\prod_{i=1}^{p}(a_i)_k x^{k+1}\right]\Big/\left[k!\prod_{i=1}^{q}[(b_i)_k\right]\\ &= x\prod_{i=1}^{p}(\theta+a_i)w.\end{aligned}$$

This shows that the ${}_pF_q$ in (13) satisfies (21) and that its specializations for $(p,q)=(3,2)$ and $(2,1)$ satisfy (12) and (11), respectively. I learned this derivation from my, now deceased, professor E. D. Rainville; see [**8**, Chapter 5].

3. Two proofs of the identity (4)–(5). First consider (15). To establish this identity one may either use the orthogonality of the Gegenbauer polynomials $C_n^\nu(x)$ with weight function $(1-x^2)^{\nu-1/2}$ or proceed as follows. On one hand, it follows easily from the generating function relationship (6) that

$$dC_n^\nu(x)/dx = 2\nu C_{n-1}^{\nu+1}(x). \tag{22}$$

If we use this relation in the expansion

$$(2x)^n/n! = \sum_{k=0}^{[n/2]} a_{k,n}(\nu)C_{n-2k}^\nu(x), \tag{23}$$

we get

$$a_{k,n}(\nu+1) = \nu a_{k,n+1}(\nu). \tag{24}$$

This relation implies that if $a_{k,n}(1)$ were known, then we could compute the values of $a_{k,n}(j)$ for all positive integers j. Since $a_{k,n}(\nu)$ must be a rational function of ν, we would then know $a_{k,n}(\nu)$ for all $\nu>0$.

On the other hand, since $C_n^1(\cos\theta)=\sin[(n+1)\theta]/\sin\theta$ and the orthogonality of the set $\{\sin(k\theta)\}$ on $(-\pi/2,\pi/2)$ is known from the elementary calculus, we have

$$\begin{aligned}\int_{-1}^{+1}C_n^1(x)C_m^1(x)(1-x^2)^{1/2}\,dx &= \int_{-\pi/2}^{+\pi/2}\sin[(n+1)\theta]\sin[(m+1)\theta]\,d\theta\\ &= (\pi/2)\,\delta_{n,m},\end{aligned}$$

and

$$\begin{aligned}&\int_{-1}^{+1}[(2x)^n/n!]C_{n-2k}^1(x)(1-x^2)^{1/2}\,dx\\ &\quad= \int_{-\pi/2}^{+\pi/2}[(2\cos\theta)^n/n!]\sin[(n-2k+1)\theta]\sin\theta\,d\theta,\\ &\quad= (\pi/2)(n-2k+1)/[k!(n-k+1)!].\end{aligned}$$

Thus, $a_{k,n}(1)=(n-k+1)/[k!(n-k+1)!]$, and the evaluation $a_{k,n}(\nu)=(n-2k+\nu)/[k!(\nu)_{n-k+1}]$ of the coefficients in (23) follows, which gives (15).

Both Hua's proof of (4)–(5) and the alternative one that follows it take advantage of (15). Using (15) in (14), we find that the coefficients c_k in (4) are given by (their

dependence on (ν, λ) is suppressed)

$$c_k = \left[(-1)^k(\nu)_{n-k}/k!\right] \sum_{i=0}^{[n/2]-k} (\lambda + n - 2k - 2i)/\left[i!(\lambda)_{n-2p-i+1}\right],$$
$$= \sum_{p+i=k} \left[(-1)^p(\nu)_{n-p}(\lambda + n - 2p - 2i)/\left[p!i!(\lambda)_{n-2k-i+1}\right],\right.$$

or

$$[\Gamma(\nu)/\Gamma(\lambda)]c_k$$
$$= [(\lambda + n - 2k)/k!] \sum_{p=0}^{k} (-1)^p \binom{k}{p} \Gamma(\nu + n - p)/\Gamma(\lambda + n - p + k + 1). \tag{25}$$

Let $\Delta g(x) = g(x+1) - g(x)$. To obtain (5) from (25) Hua uses the iterated finite difference formula

$$\Delta^k[g(a+x)/g(b+x)] = \sum_{p=0}^{k} (-1)^p \binom{k}{p} g(a + x + k - p)/g(b + x + k - p),$$

with $x = n$, $a = \nu - k$ and $b = \lambda - 2k + 1$. Then the sum in (25) becomes

$$\sum_{p=0}^{k} \left[(-1)^p \binom{k}{p} [\Gamma(\nu + n - p)/\Gamma(\lambda + n - p - k + 1)]\right.$$
$$= \Delta^k[\Gamma(\nu + n - k)/\Gamma(\lambda + n - 2k + 1). \tag{26}$$

The functional equation for the gamma function, $x\Gamma(x) = \Gamma(x+1)$, and some algebra yield the identity

$$\Delta[\Gamma(a+n)/\Gamma(b+n)] = (a-b)\Gamma(a+n)/\Gamma(b+n+1).$$

This result and the functional equation for Γ imply in turn that

$$\Delta^q[\Gamma(\alpha+n)/\Gamma(b+n)]$$
$$= \Gamma(a-b+1)\Gamma(a+n)/\Gamma(a-b-q+1)\Gamma(b+n+q). \tag{27}$$

Finally, if we use (27) with $q = k$, $a = \nu - k$ and $b = \lambda - 2k + 1$, we conclude that the right-hand side of (27) becomes

$$\Gamma(\nu - \lambda + k)\Gamma(\nu + n - k)/\Gamma(\nu - \lambda)\Gamma(n + \lambda - k + 1),$$

which implies (5).

I now give a second proof of (4)–(5) modeled on arguments of Rainville [**8**, p. 179–181]. It is short and sweet. However, it does depend upon the formula ${}_2F_1(-n, a; c; 1) = (c-a)_n/(c)_n$ [**8**, p. 69], which I do not derive here. Using it,

one may rewrite the sum in (26) as

$$\sum_{p=0}^{k} [(-k)_p\Gamma(\nu+n-p)/[p!\Gamma(\lambda+n-p-k+1)]$$
$$= [\Gamma(\nu+n)/\Gamma(\lambda+n-k+1)]$$
$$\times \sum_{p=0}^{\infty} (-k)_p(k-n-\lambda)_p/[p!(1-n-\nu)_p],$$
$$= [\Gamma(\nu+n)/\Gamma(\lambda+n-k+1)]_2F_1(-k, k-\lambda-n; 1-\nu-n; 1);$$
$$= [\Gamma(\nu+n)/\Gamma(\lambda+n-k+1)](1+\lambda-k-\nu)_k/(1+\nu-n)_k,$$

which reduces to $\Gamma(\nu-\lambda+k)\Gamma(\nu+n-k)/\Gamma(\nu-\lambda)\Gamma(n+\lambda-k+1)$.

4. Askey and Gasper's proof of (3). Askey and Gasper actually proved that

$$S(x, \alpha, n) \equiv \sum_{k=0}^{n} P_k^{(\alpha,0)}(x) > 0$$

for $-1 \leqslant x \leqslant 1$ and $\alpha > -2$ in [**1**, Theorem 3], where the $P_k^{(\alpha,\beta)}(x)$ are Jacobi polynomials and the sum $S(x, \alpha, n)$ is a constant multiple of $F(x, \alpha, n)$. However, for our purposes it suffices to establish the inequality (2), for $\alpha > -1$ and $-1 < x \leqslant 1$, which we proceed to do.

The operator $T_{e,c}$, implicitly used by Askey and Gasper, as applied to t^n is

$$T_{e,c}t^k = \{\Gamma(e)t^{1-e}/\Gamma(c)\Gamma(e-c)\}\int_0^t (t-y)^{e-c-1}y^{c+k-1}\,dy. \tag{28}$$

If we let $y = ts$ in (28), note that for the beta function

$$B(\alpha, \beta) = \int_0^1 (1-s)^{\alpha-1}s^{\beta-1}\,ds = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$$

and use the obvious consequence $\Gamma(a+n)/\Gamma(a) = (a)_n$ of the functional equation for the gamma function, then we find that

$$T_{e,c}t^k = (c)_k t^k/(e)k;$$

hence, (10) holds.

We now prove (2). First, observe, that (2) is not trivial; (3) cannot be directly applied to the ${}_3F_2$ in (2) to give the desired inequality. However if we use the result (8)–(9) and the transformation $T_{e,c}$ with properly chosen e, and c, we can apply (3) to obtain (2). By (2) and (10),

$$F(x, \alpha, n) = T_{\alpha+1,(1/2)(\alpha+1)}\ {}_2F_1\big(-n, -n+\alpha+2; \alpha+\tfrac{1}{2}; t\big), \tag{29}$$

where $t = \frac{1}{2}(1-x)$. We apply (8) to (29) with $\nu = \frac{1}{2}(\alpha+2)$; and $\lambda = \frac{1}{2}(\alpha+1)$. The result is

$$F(x, \alpha, n) = T_{\alpha+1,(1/2)(\alpha+1)} \sum_{j=0}^{[n/2]} d_j\big(\tfrac{1}{2}(\alpha+2), \tfrac{1}{2}(\alpha+1)\big)$$
$$\times {}_2F_1\big(2j-n, n-2j+\alpha+1; \tfrac{1}{2}(\alpha+2); t\big);$$

or, since we may reverse the order of $T_{\alpha+1,(\alpha+1)/2}$ and the summation

$$F(x, \alpha, n) = \sum_{j=0}^{[n/2]} d_j T_{\alpha+1,(1/2)(\alpha+1)}\ {}_2F_1\big(2j-n, n-2j+\alpha+1; \tfrac{1}{2}(\alpha+2); t\big), \tag{30}$$

where we have suppressed the dependence of the d_j on α. We next evaluate $T_{\alpha+1,(\alpha+1)/2}$ acting on the ${}_2F_1$'s in (30) using (10). The result is:

$$F(x, \alpha, n) = \sum_{j=0}^{[n/2]} d_j\ \ {}_3F_2\big(2j-n, n-2j+\alpha+1, \tfrac{1}{2}(\alpha+1); \alpha+1, \tfrac{1}{2}(\alpha+2); t\big).$$

But these ${}_3F_2$'s now precisely fit (3) with $2a = 2j - n$ and $2b = n - 2j + \alpha + 1$. This comes about since we had a free choice of λ in $(0, \nu)$ in (8)! Thus using (3) we find that

$$F(x, \alpha, n) = \sum_{j=0}^{[n/2]} d_j\big[{}_2F_1\big(\tfrac{1}{2}(2j-n), \tfrac{1}{2}(n-2j+\alpha+1); \tfrac{1}{2}(\alpha+2); t\big)\big]^2, \tag{31}$$

By (5) and (9), the coefficients $d_j(\frac{1}{2}(\alpha+2), \frac{1}{2}(\alpha+1))$ in (31) are all positive for $\alpha > -1$. At least one of the ${}_2F_1$'s in (31) is always positive. To see this we observe that if n is even, then the ${}_2F_1$ in (31) corresponding to $j = [n/2]$ is (recall that $t = \frac{1}{2}(1-x)$)

$${}_2F_1\big(0, \tfrac{1}{2}(\alpha+1); \tfrac{1}{2}(\alpha+2); \tfrac{1}{2}(1-x)\big) \equiv 1 > 0;$$

and if n is odd, then the ${}_2F_1$ in (26) corresponding to $j = [n/2]$ is

$${}_2F_1\big(-\tfrac{1}{2}, \tfrac{1}{2}(\alpha+2); \tfrac{1}{2}(\alpha+2); \tfrac{1}{2}(1-x)\big) \equiv [(1+x)/2]^{1/2} > 0$$

for $-1 < x \leqslant 1$. This completes the proof of (2).

REFERENCES

1. R. Askey and G. Gasper, Positive Jacobi polynomial sums II, *Amer. J. Math.*, 98 (1976) 709–737.
2. Th. Clausen, Ueber die Fälle, wenn die Reihe von der Form $y = 1 + \frac{\alpha\beta}{\gamma}x + \frac{\alpha(\alpha+1)\beta(\beta+1)}{\gamma(\gamma+1)2!}x^2 +$ etc. ein Quadrat von der Form $z = 1 + \frac{\alpha'\beta'\delta'}{\gamma'\varepsilon'}x + \frac{\alpha'(\alpha'+1)\beta'(\beta'+1)\delta'(\delta'+1)}{\gamma'(\gamma'+1)\varepsilon'(\varepsilon'+1)2!}x^2 +$ etc. hat. *J. für die reine und angewandte Math*, 3 (1828) 89–91.
3. Th. Clausen, Beitrag zur Theorie der Reihen, *J. für die reine und angewandte Math.*, 3 (1828) 92–95.
4. L. de Branges, A proof of the Bieberbach conjecture, *Acta Math*, 154 (1985) 137–152.
5. L. K. Hua, *Harmonic Analysis of Functions of Several Complex Variables in the Classical Domains*, Trans. of Math. Monographs, Amer. Math. Soc. 6 (1963), Providence, R.I.
6. J. Korevaar, Ludwig Bieberbach's Conjecture and its Proof, this MONTHLY, 93 (1986) 505–513.
7. W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics*, Springer-Verlag, N.Y., 1966.
8. E. D. Rainville, *Special Functions*, Macmillan Co., New York, 1960.

The Strong Law of Small Numbers

RICHARD K. GUY
Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4

This article is in two parts, the first of which is a do-it-yourself operation, in which I'll show you 35 examples of patterns that *seem* to appear when we look at several small values of n, in various problems whose answers depend on n. The question will be, in each case: do you think that the pattern persists for all n, or do you believe that it is a figment of the smallness of the values of n that are worked out in the examples?

Caution: examples of both kinds appear; they are not all figments!

In the second part I'll give you the answers, insofar as I know them, together with references.

Try keeping a scorecard: for each example, enter your opinion as to whether the observed pattern is known to continue, known not to continue, or not known at all.

This first part contains no information; rather it contains a good deal of disinformation. The first part contains one theorem:

$\boxed{\text{You can't tell by looking.}}$

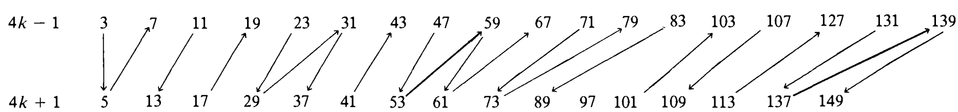
It has wide application, outside mathematics as well as within. It will be proved by intimidation.

Here are some well-known examples to get you started.

Example 1. The numbers $2^{2^0}+1=3$, $2^{2^1}+1=5$, $2^{2^2}+1=17$, $2^{2^3}+1=257$, $2^{2^4}+1=65537$, are primes.

Example 2. The number 2^n-1 can't be prime unless n is prime, but $2^2-1=3$, $2^3-1=7$, $2^5-1=31$, $2^7-1=127$, are primes.

Example 3. Apart from 2, the oddest prime, all primes are either of shape $4k-1$, or of shape $4k+1$. In any interval $[1, n]$, the former are at least as numerous as the latter ($4k-1$ wins the "prime number race"):



Example 4. Pick several numbers at random (it suffices just to look at odd ones). Estimate the probability that a number has more divisors of shape $4k-1$, than it does of shape $4k+1$. For example, 21 has two of the first kind (3 & 7) and two of the second (1 & 21), while 25 has all three $(1, 5, 25)$ of the second kind.

Example 5. The five circles of FIG. 1 have $n = 1, 2, 3, 4, 5$ points on them. These points are in general position, in the sense that no three of the $\binom{n}{2}$ chords joining

them are concurrent. Count the numbers of regions into which the chords partition each circle.

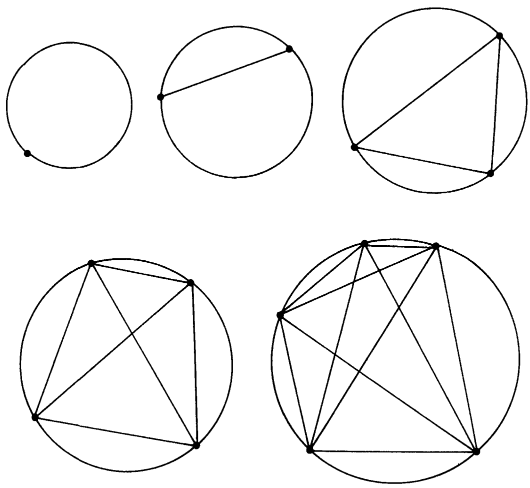


FIG. 1. How many regions in each of these circles?

I've been trying to formulate the **Strong Law of Small Numbers** for many years [**9**]. The best I can do so far is

There aren't enough small numbers to meet the many demands made of them.

It is the enemy of mathematical discovery. When you notice a mathematical pattern, how do you know it's for real?

Superficial similarities spawn spurious statements.

Capricious coincidences cause careless conjectures.

On the other hand, the Strong Law often works the other way:

Early exceptions eclipse eventual essentials.

Initial irregularities inhibit incisive intuition.

Here are some misleading facts about small numbers:

Ten per cent of the first hundred numbers are perfect squares.

A quarter of the numbers less than 100 are primes.

Except for 6, all numbers less than 10 are prime powers.

Half the numbers less than 10 are Fibonacci numbers

$$0, 1, 1, 2, 3, 5, 8, \dots$$

and alternate Fibonacci numbers, $1, 2, 5, \dots$ are both Bell numbers and Catalan numbers.

Example 6. The numbers 31, 331, 3331, 33331, 333331, 3333331, are each prime.

Example 7. The alternating sums of factorials,

$$\begin{gathered} 3! - 2! + 1! = 5 \\ 4! - 3! + 2! - 1! = 19 \\ 5! - 4! + 3! - 2! + 1! = 101 \\ 6! - 5! + 4! - 3! + 2! - 1! = 619 \\ 7! - 6! + 5! - 4! + 3! - 2! + 1! = 4421 \\ 8! - 7! + 6! - 5! + 4! - 3! + 2! - 1! = 35899 \end{gathered}$$

are each prime.

Example 8. In the table

row 1	1 1	2
row 2	1 2 1	3
row 3	1 3 2 3 1	5
row 4	1 4 3 2 3 4 1	7
row 5	1 5 4 3 5 2 5 3 4 5 1	11
row 6	1 6 5 4 3 5 2 5 3 4 5 6 1	13
row 7	1 7 6 5 4 7 3 5 7 2 7 5 3 7 4 5 6 7 1	19
row 8	1 8 7 6 5 4 7 3 8 5 7 2 7 5 8 3 7 4 5 6 7 8 1	23
row 9	1 9 8 7 6 5 9 4 7 3 8 5 7 9 2 9 7 5 8 3 7 4 9 5 6 7 8 9 1	29

row n is obtained from row $n - 1$ by inserting n between each pair of consecutive numbers which add to n. The number of numbers in each row is shown on the right. Each is prime.

Example 9. Is there a prime of shape $7013 \times 2^n + 1$?

Example 10. Are all the numbers $78557 \times 2^n + 1$ composite?

Example 11. When you use Euclid's method to show that there are unboundedly many primes:

$$\begin{gathered} 2 + 1 = 3 \\ (2 \times 3) + 1 = 7 \\ (2 \times 3 \times 5) + 1 = 31 \\ (2 \times 3 \times 5 \times 7) + 1 = 211 \\ (2 \times 3 \times 5 \times 7 \times 11) + 1 = 2311 \end{gathered}$$

you don't always get primes:

$$(2 \times 3 \times 5 \times 7 \times 11 \times 13) + 1 = 30031 = 59 \times 509$$
$$(2 \times 3 \times 5 \times 7 \times 11 \times 13 \times 17) + 1 = 510511 = 19 \times 97 \times 277$$
$$(2 \times 3 \times 5 \times 7 \times 11 \times 13 \times 17 \times 19) + 1 = 9699691 = 347 \times 27953$$

but if you go to the *next* prime, its difference from the product is always a prime:

$$5 - 2 = 3$$
$$11 - (2 \times 3) = 5$$
$$37 - (2 \times 3 \times 5) = 7$$
$$223 - (2 \times 3 \times 5 \times 7) = 13$$
$$2333 - (2 \times 3 \times 5 \times 7 \times 11) = 23$$
$$30047 - (2 \times 3 \times 5 \times 7 \times 11 \times 13) = 17$$
$$510529 - (2 \times 3 \times 5 \times 7 \times 11 \times 13 \times 17) = 19$$
$$9699713 - (2 \times 3 \times 5 \times 7 \times 11 \times 13 \times 17 \times 19) = 23$$

Example 12. From the sequence of primes, form the first differences, then the absolute values of the second, third, fourth,... differences:

```
2 3 5 7 11 13 17 19 23 29 31 37 41 43 47 53 59 61 67
 1 2 2 4 2 4 2 4 6 2 6 4 2 4 6 6 2 6
  1 0 2 2 2 2 2 2 4 4 2 2 2 2 0 4 4 2
   1 2 0 0 0 0 0 2 0 2 0 0 0 2 4 0 2
    1 2 0 0 0 0 2 2 2 2 0 0 2 2 4 2 2
     1 2 0 0 0 2 0 0 0 2 0 2 0 2 2 0
      1 2 0 0 2 2 0 0 2 2 2 2 2 0 2 0
       1 2 0 2 0 2 0 2 0 0 0 0 2 2 2
        1 2 2 2 2 2 2 2 0 0 0 2 0 0 0
         1 0 0 0 0 0 0 2 0 0 2 2 0 0
          1 0 0 0 0 0 2 2 0 2 0 2 0 0
           1 0 0 0 0 2 0 2 2 2 2 2 0
```

Is the first term in each sequence of differences always 1?

Example 13. 2^n is never congruent to 1 (mod n) for $n > 1$. 2^n is congruent to 2 (mod n) whenever n is prime, and occasionally when it isn't ($n = 341, 561, \ldots$). Is 2^n ever congruent to 3 (mod n) for $n > 1$?

Example 14. The good approximations to $5^{1/5}$, namely, the convergents to

$$1 + \frac{1}{2+}\,\frac{1}{1+}\,\frac{1}{1+}\,\frac{1}{1+}\,\frac{1}{2+}\cdots \text{ are } \frac{1}{1}, \frac{3}{2}, \frac{4}{3}, \frac{7}{5}, \frac{11}{8}, \frac{29}{21}, \ldots$$

which have Fibonacci numbers for denominators and Lucas numbers for numerators.

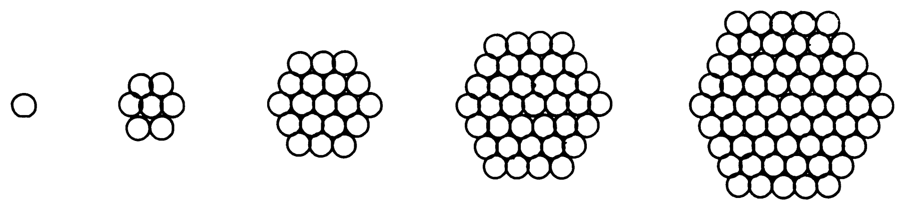
Example 15.

$$(x+y)^3 = x^3 + y^3 + 3xy(x+y)(x^2+xy+y^2)^0$$

$$(x+y)^5 = x^5 + y^5 + 5xy(x+y)(x^2+xy+y^2)^1$$

$$(x+y)^7 = x^7 + y^7 + 7xy(x+y)(x^2+xy+y^2)^2$$

Example 16. The sequence of **hex numbers** (so named to distinguish them from the **hexagonal** numbers, $n(2n-1)$) are depicted in FIG. 2.
The partial sums of this sequence, $1, 8, 27, 64, 125$, appear to be perfect cubes.



1 7 19 37 61

FIG. 2. The hex numbers.

Example 17. Write down the positive integers, delete every second, and form the partial sums of those remaining:

1	$\cancel{2}$	3	$\cancel{4}$	5	$\cancel{6}$	7	$\cancel{8}$	9	$\cancel{10}$	11
1		4		9		16		25		36

Example 18. As before, but delete every third, then delete every second partial sum:

1	2	$\cancel{3}$	4	5	$\cancel{6}$	7	8	$\cancel{9}$	10	11	$\cancel{12}$	13	14	$\cancel{15}$	16
1	$\cancel{3}$		7	$\cancel{12}$		19	$\cancel{27}$		37	$\cancel{48}$		61	$\cancel{75}$		91
1			8			27			64			125			216

Example 19. Again, but delete every fourth, then every third partial sum, then every second of their partial sums:

1	2	3	$\cancel{4}$	5	6	7	$\cancel{8}$	9	10	11	$\cancel{12}$	13	14	15	$\cancel{16}$	17
1	3	$\cancel{6}$		11	17	$\cancel{24}$		33	43	$\cancel{54}$		67	81	$\cancel{96}$		113
1	$\cancel{4}$			15	$\cancel{32}$			65	$\cancel{108}$			175	$\cancel{256}$			369
1				16				81				256				625

Example 20. Again, but circle the first number of the sequence, delete the second after that, the third after that, and so on. Form the partial sums and repeat:

1 2 ~~3~~ 4 5 ~~6~~ 7 8 9 ~~10~~ 11 12 13 14 ~~15~~ 16 17 18 19 20 ~~21~~

2 6 ~~11~~ 18 26 ~~35~~ 46 58 71 ~~85~~ 101 118 136 155 ~~175~~

6 24 ~~50~~ 96 154 ~~225~~ 326 444 580 ~~735~~

24 120 ~~274~~ 600 1044 ~~1624~~

120 720 ~~1764~~

720

Example 21. Write down the odd numbers starting with 43. Circle 43, delete one number, circle 47, delete two numbers, circle 53, delete three numbers, circle 61, and so on. The circled numbers are prime (FIG. 3)

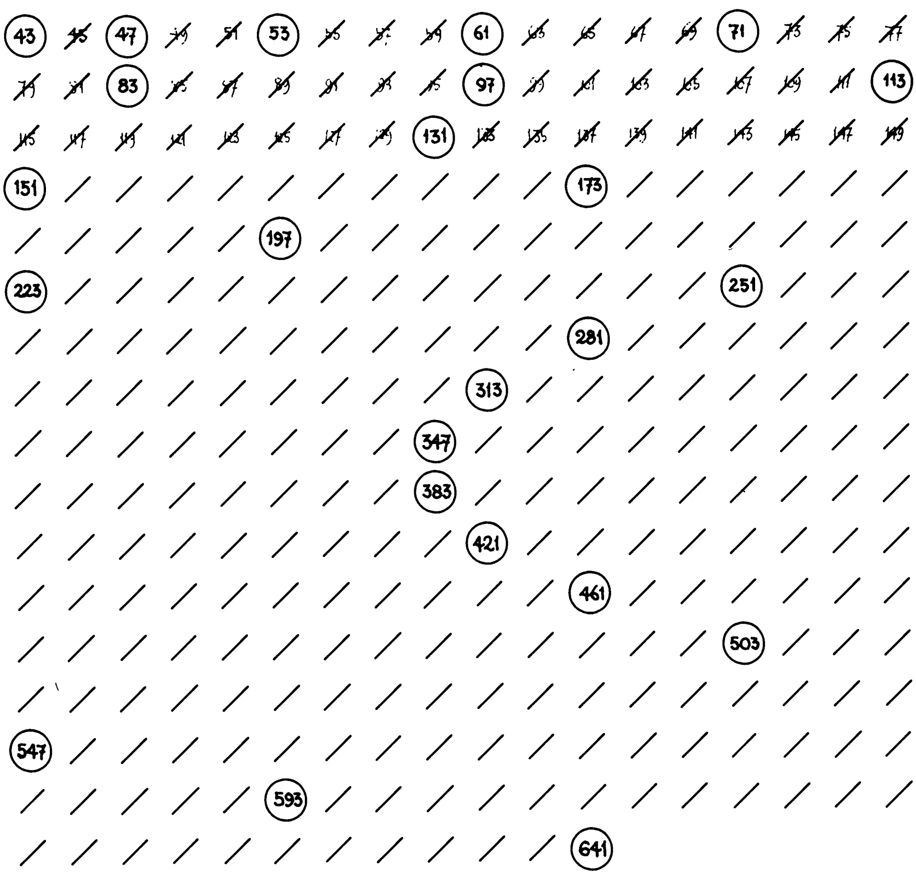


FIG. 3. Parabolas of primes remain.

Example 22. In Table 1 the odd prime values of $n^4 + 1$ and of $17 \times 2^n - 1$ are printed in **bold**. They occur simultaneously for $n = 2, 4, 6, 16, 20$.

TABLE 1

n	$n^4 + 1$	$17 \times 2^n - 1$
0	1	$16 = 2^4$
1	2	$33 = 3 \times 11$
2	**17**	**67**
3	$82 = 2 \times 41$	$135 = 3^3 \times$
4	**257**	**271**
5	$626 = 2 \times$	$543 = 3 \times$
6	**1297**	**1087**
7	$2402 = 2 \times$	$2175 = 3 \times$
8	$4097 = 17 \times$	$4351 = 19 \times$
9	$6562 = 2 \times$	$8703 = 3^2 \times$
10	$10001 = 73 \times$	$17407 = 13^2 \times$
11	$14642 = 2 \times$	$34815 = 3 \times$
12	$20737 = 89 \times$	$69631 = 179 \times$
13	$28562 = 2 \times$	$139263 = 3 \times$
14	$38417 = 41 \times$	$278527 = 223 \times$
15	$50626 = 2 \times$	$557055 = 3^2 \times$
16	**65537**	**1114111**
17	$83522 = 2 \times$	$2228223 = 3 \times$
18	$104977 = 113 \times$	$4456447 = 59 \times$
19	$130322 = 2 \times$	$8912895 = 3 \times$
20	**160001**	**17825791**
21	$194482 = 2 \times$	$35651583 = 3^4 \times$
22	$234257 = 73 \times$	$71303167 = 13 \times$
23	$279842 = 2 \times$	$142606335 = 3 \times$

Example 23. In Table 2 the prime values of $21 \times 2^n - 1$ and of $7 \times 4^n + 1$ are printed in **bold**. They occur simultaneously for $n = 1, 2, 3, 7, 10, 13$.

TABLE 2

n	$21 \times 2^n - 1$	$7 \times 4^n + 1$
0	$20 = 2^2 \times 5$	$8 = 2^3$
1	**41**	**29**
2	**83**	**113**
3	**167**	**449**
4	$335 = 5 \times$	$1793 = 11 \times$
5	$671 = 11 \times$	$7169 = 67 \times$
6	$1343 = 17 \times$	$28673 = 53 \times$
7	**2687**	**114689**
8	$5375 = 5^3 \times$	$458753 = 79 \times$
9	$10751 = 13 \times$	$1835009 = 11 \times$
10	**21503**	**7340033**
11	$43007 = 29 \times$	$29360129 = 37 \times$
12	$86015 = 5 \times$	$117440513 = 3907 \times$
13	**172031**	**469762049**
14	$344063 = 17 \times$	$1879048193 = 11 \times$
15	$688127 = 11^4 \times$	$7516192769 = 29^2 \times$
16	$1376255 = 5 \times$	$30064771073 = 113 \times$
17	$2752511 = 19 \times$	$120259084289 = 379 \times$

Example 24. Consider the sequence

$$x_0 = 1, \qquad x_{n+1} = \left(1 + x_0^2 + x_1^2 + \cdots + x_n^2\right)/(n+1) \qquad (n \geqslant 0).$$

n	0	1	2	3	4	5	6	7	8	9	...
x_n	1	2	3	5	10	28	154	3520	1551880	267593772160	...

Is x_n always an integer?

Example 25. The same, but with cubes in place of squares: $y_0 = 1$, $y_{n+1} = (1 + y_0^3 + y_1^3 + \cdots + y_n^3)/(n+1)$ $(n \geqslant 0)$. Same question.

n	0	1	2	3	4	5	...
y_n	1	2	5	45	22815	2375152056927	...

Example 26. Also for fourth powers, $z_{n+1} = (1 + z_0^4 + z_1^4 + \cdots + z_n^4)/(n+1)$.

n	0	1	2	3	4	...
z_n	1	2	9	2193	5782218987645	...

And for fifth powers, and so on.

Example 27. The irreducible factors of $x^n - 1$ are **cyclotomic polynomials**, i.e., $x^n - 1 = \prod_{d|n}\Phi_d(x)$, so that $\Phi_1(x) = x - 1$, $\Phi_2(x) = x + 1$, $\Phi_3(x) = x^2 + x + 1$, $\Phi_4(x) = x^2 + 1$. The cyclotomic polynomial of order n, $\Phi_n(x)$, has degree $\varphi(n)$, Euler's totient function. It is easy to write down $\Phi_n(x)$ if n is prime, twice a prime, or a power of a prime, and for many other cases. Are the coefficients always ± 1 or 0?

Example 28. If two people play Beans-Don't-Talk, the typical position is a whole number, n, and there are just two options, from n to $(3n \pm 1)/2^*$, where 2^* means the highest power of 2 that divides the numerator. The winner is the player who moves to 1. For example, 7 is a **$\mathscr{P}$-position**, a previous-player-winning position, because the opponent must go to

$$(3 \times 7 + 1)/2 = 11 \text{ or } (3 \times 7 - 1)/2^2 = 5$$

and 11 and 5 are **$\mathscr{N}$-positions**, next-player-winning positions, since they have the options $(3 \times 11 - 1)/2^5 = 1$ and $(3 \times 5 + 1)/2^4 = 1$.

If τ is the probability that a number is an $\mathscr{N}$-position, and there are no $\mathscr{O}$-positions (from which neither player can force a win), then the probability that a number is a $\mathscr{P}$-position is $1 - \tau$. This happens just if both options are $\mathscr{N}$-positions, so $1 - \tau = \tau^2$, and τ is the golden ratio, $(\sqrt{5} - 1)/2 \approx 0 \cdot 618$.

So it is no surprise that 5 out of the first 8 numbers are $\mathscr{N}$-positions, 8 out of the first 13, 13 of the first 21, 21 of the first 34, and 34 of the first 55, since the ratio of consecutive Fibonacci numbers tends to the golden ratio.

Example 29. Does each of the two diophantine equations

$$2x^2(x^2 - 1) = 3(y^2 - 1) \text{ and } x(x-1)/2 = 2^n - 1$$

have just the five positive solutions $x = 1, 2, 3, 6,$ and 91?

Example 30. Consider the sequence $a_1 = 1$, $a_{n+1} = \left\lfloor \sqrt{2a_n(a_n+1)} \right\rfloor$ $(n \geqslant 1)$

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
a_n	1	2	3	4	6	9	13	19	27	38	54	77	109	154	218	309	437	618	874	1236	1748
			1		2		4		8		16		32		64		128		256		512

Are alternate differences, $a_{2k+1} - a_{2k}$, the powers of two, 2^k?

Example 31. In the same sequence, are the even ranked members, a_{2k+2}, given by $2a_{2k} + \varepsilon_k$, where ε_k is the kth digit in the binary expansion of $\sqrt{2} = 1.01101010000010\ldots$?

Example 32. Is this the same sequence as $a_1 = 1$, $a_2 = 2$, $a_3 = 3$, $a_{n+1} = a_n + a_{n-2}$ $(n \geqslant 3)$?

Example 33. The nth derivative of x^x, evaluated at $x = 1$, is an integer. Is it always a multiple of n? Values for $n = 1, 2, 3, \ldots$ are

$$1 \times 1, 2 \times 1, 3 \times 1, 4 \times 2, 5 \times 2, 6 \times 9, 7 \times (-6), 8 \times 118, 9 \times (-568),$$
$$10 \times 4716, 11 \times (-38160), 12 \times 358126, 13 \times (-3662088), 14 \times 41073096,$$
$$15 \times (-500013528), 16 \times 6573808200, 17 \times (-92840971200),$$
$$18 \times 1402148010528, \ldots .$$

Example 34. In how many ways, c_n, can you arrange n pennies in rows, where every penny in a row above the first must touch two adjacent pennies in the row below?

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
c_n	1	1	1	2	3	5	9	15	26	45	78	135	234	406	704	1222	2120

To throw more light on such sequences, partition theorists often express their generating function

$$\sum_{n=0}^{\infty} c_n x^n = 1 + x + x^2 + 2x^3 + 3x^4 + 5x^5 + 9x^6 + 15x^7 + \cdots$$

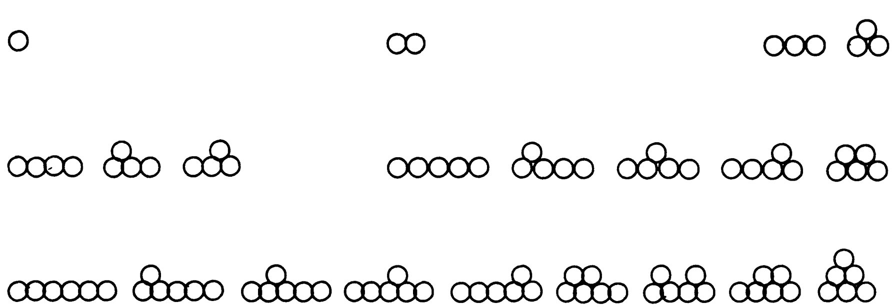


FIG. 4. Propp's penny partitions.

as an infinite product,

$$\prod_{n=1}^{\infty}(1-x^n)^{-a(n)}$$

In this case, $a(n)$ are consecutive Fibonacci numbers:

n	1	2	3	4	5	6	7	8	9	10	...
$a(n)$	1	0	1	1	2	3	5	8	13	21	...

Example 35. If p_k is the kth prime, $p_1 = 2$, $p_2 = 3, \ldots,$ does

$$\prod_{k=1}^{\infty}(1-x^{p_k})^{-1} = 1 + \sum_{k=1}^{\infty}\frac{x^{p_1+p_2+\cdots+p_k}}{(1-x)(1-x^2)\cdots(1-x^k)}?$$

Answers

1. No less a person than Fermat was fooled by the Strong Law! Euler gave the factorization $2^{32}+1 = 641 \times 6700417$. All other known examples of Fermat numbers are composite; Jeff Young & Duncan Buell **[32]** have recently shown that $2^{2^{20}}+1$ is composite.

2. There are very few **Mersenne primes**, $2^p - 1$. No one can prove that there are infinitely many; $2^{11} - 1 = 23 \times 89$ is not one. See A3 in **[12]** and sequence 1080 in **[28]**.

3. In the "prime number race," $4k-1$ and $4k+1$ alternately take the lead infinitely often. This was proved by Littlewood **[18]**. For many papers on this subject see N–12 of *Reviews in Number Theory*, for example, Chen **[4]**.

4. A theorem of Legendre (see **[6]**, for example) states that if D_+ and D_- are the numbers of divisors of n of shapes $4k+1$ and $4k-1$, then the number of representations of n as the sum of two squares is $4(D_+ - D_-)$. So $D_+ \geqslant D_-$ for every number!

5. Before we reveal all, here is a circle (FIG. 5) with ten points to further confuse you. It has 256 regions.

If the circle has n points, there are $\binom{n}{4}$ intersections of chords inside the circle, since each set of four points gives just one such intersection. The number of vertices in the figure is $V = n + \binom{n}{4}$. To find the number of edges, count their ends. There are $n+1$ at each of the n points and four at each of the $\binom{n}{4}$ intersections, so $2\mathrm{E} = n(n+1) + 4\binom{n}{4}$. By Euler's formula, the number of regions inside the circle is

$$\begin{aligned}
\mathrm{E} + 1 - V &= 2\binom{n}{4} + \frac{1}{2}n(n+1) + 1 - \binom{n}{4} - n \\
&= \binom{n}{4} + \frac{1}{2}n(n-1) + 1 \\
&= \binom{n-1}{4} + \binom{n-1}{3} + \binom{n-1}{2} + \binom{n-1}{1} + \binom{n-1}{0}.
\end{aligned}$$

A direct proof, by labelling the regions with at most four of the numbers $1, 2, \ldots,$

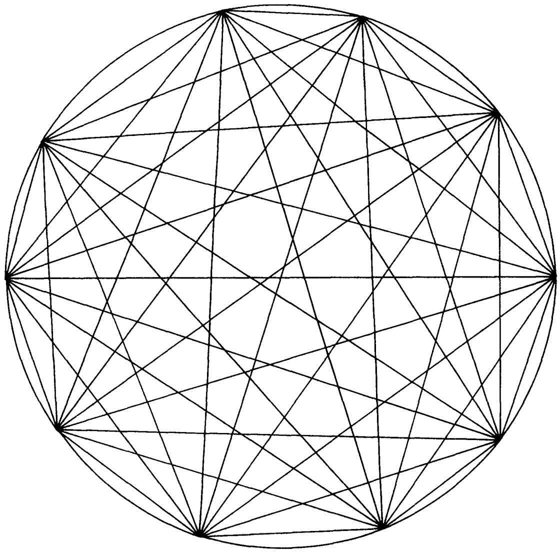


FIG. 5. Circle partitioned into 2^8 regions.

$n - 1$, will appear in [**5**]. The answer is just five of the n terms in the binomial expansion of $(1 + 1)^{n-1}$. For $n < 6$, this is all the terms, and the number is a power of 2. For $n = 6$, only 1 is missing. For $n = 10$ just half the terms are missing, and the number of regions is $\frac{1}{2} \cdot 2^9 = 256$.

# of points =	1	2	3	4	5	6	7	8	9	10	11	12	13	14
# of regions =	**1**	**2**	**4**	**8**	**16**	31	57	99	163	**256**	386	562	794	1093

Some other famous numbers, e.g. 163 and 1093, also occur in this sequence, number 427 in [**28**].

6. No member of this sequence is divisible by 2, 3, 5, 7, 11, 13, or 37, as may be seen immediately from well known divisibility tests. On the other hand, 17, 19, 23, 29, 31,... divide 33...331 just if the number of threes is respectively $16k + 8$, $18k + 11$, $22k + 20$, $28k + 19$, $15k + 1, \ldots,$ while 41, 43, 53, 67, 71, 73, 79,... divide no members of the sequence. I don't think that there is a simple description of which primes do, and which primes don't, divide. The next member, 33333331, is also prime, but $333333331 = 17 \times 19607843$

7. We've again given ourselves a good start, since $\sum_{k=1}^{n}(-1)^{n-k}k!$ is not divisible by any prime $\leqslant n$. However,

$$9! - 8! + 7! - 6! + 5! - 4! + 3! - 2! + 1! = 326981 = 79 \times 4139.$$

8. This example, as well as example **5.**, was first shown to me by Leo Moser, a quarter of a century ago. Row n is the list of denominators of the **Farey series** of order n, i.e., the set of rational fractions r, $0 \leqslant r \leqslant 1$, whose denominators do not exceed n. In getting row n from row $n - 1$, just $\varphi(n)$ numbers are inserted, where $\varphi(n)$ is Euler's totient function, the number of numbers not exceeding n which are prime to n. It is fortuitous that $1 + \sum_{k=1}^{n}\varphi(k)$ is prime for $1 \leqslant n \leqslant 9$. As $\varphi(10) = 4$, the number of numbers in row 10 is $29 + 4 = 33$, and is not prime.

9. The expression $7013 \times 2^n + 1$ is composite for $0 \leqslant n \leqslant 24160$ [**15**]. Duncan Buell & Jeff Young have sieved out 325 further candidates $n < 10^5$ which might yield a prime. None is known, though it's likely that there is one.
10. The number $78557 \times 2^n + 1$ is always divisible by at least one of $3, 5, 7, 13, 19, 37, 73$ [**26, 27**]. For this and the previous example, see also B21 in [**12**].
11. R. F. Fortune conjectured that these differences are always prime: see [**8**], [**9**] and A2 in [**12**]. The next few are $37, 61, 67, 61, 71, 47, 107, 59, 61, 109, 89, 103, 79$. There's a high probability that the conjecture is true, because the difference can't be divisible by any of the first k primes, so the smallest composite candidate for $P = \prod p_k$ is p_{k+1}^2, which is approximately $(k \ln k)^2$ in size. The product of the first k primes is about e^k: to find a counter example we need a gap in the primes near N of size at least $(\ln N \operatorname{lnln} N)^2$. Such gaps are believed not to exist, but it's beyond our present means to prove this.
12. This is N. L. Gilbreath's conjecture, which has been verified for $k < 63419$ [**16**]. Hallard Croft has suggested that it has nothing to do with primes as such, but will be true for any sequence consisting of 2 and odd numbers, which doesn't increase too fast, or have too large gaps: A10 in [**12**]. In an 87-08-03 letter, Andy Odlyzko reported that he had verified the conjecture for $k < 10^{10}$.
13. D. H. & Emma Lehmer discovered that $2^n \equiv 3 \pmod{n}$ for $n = 4700063497$, but for no smaller $n > 1$.
14. The kth Lucas number and the $(k+1)$th Fibonacci number are

$$\left(\frac{1+\sqrt{5}}{2}\right)^k + \left(\frac{1-\sqrt{5}}{2}\right)^k \text{ and } \frac{1}{\sqrt{5}}\left\{\left(\frac{1+\sqrt{5}}{2}\right)^{k+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{k+1}\right\}.$$

Their ratio, as k gets large, approaches $(5-\sqrt{5})/2 \approx 1 \cdot 381966011$, whereas $5^{1/5} \approx 1 \cdot 379729661$. The next few convergents to $5^{1/5}$,

$$\frac{40}{29}, \frac{109}{79}, \frac{912}{661}, \frac{1021}{740}, \frac{26437}{19161}, \frac{27458}{19901},$$

do not involve Fibonacci or Lucas numbers. Compare sequences 256 & 260 and 924 & 925 in [**28**]. This example goes back to 1866 [**25**].
15. This is quite fortuitous [**30**]. Put $x = y = 1$, giving $2^{2n+1} - 2 = (2n+1) \times 2 \times 3^{n-1}$. It's true that

$$2^2 - 1 = 3 \times 3^0, \qquad 2^4 - 1 = 5 \times 3^1, \qquad 2^6 - 1 = 7 \times 3^2$$

but it's clear that the pattern can't continue.
16. The $(n+1)$th hex number, $1 + 6 + 12 + \cdots + 6n = 3n^2 + 3n + 1$, when added to n^3, gives $(n+1)^3$, so the pattern is genuine. It is instructive to regard the nth hex number as comprising the three faces at one corner of a cubic stack of n^3 unit cubes (FIG. 6).
17, **18**, **19**, and **20** are examples of **Moessner's process**, which does indeed produce the square, cubes, fourth powers and factorials. Moessner's paper [**20**] is followed by a proof by Perron. Subsequent generalizations are due to Paasche [**22**]: see [**19**] for a more recent exposition.
21. A thinly disguised arrangement of Euler's formula, $n^2 + n + 41$, which gives primes for $-40 \leqslant n \leqslant 39$. For $n = 40$, $n^2 + n + 41 = 41^2$. See A1 and Fig. 1 in [**12**]. For remarkable connexions with quadratic fields, continued fractions, modular functions and class numbers, see [**29**].

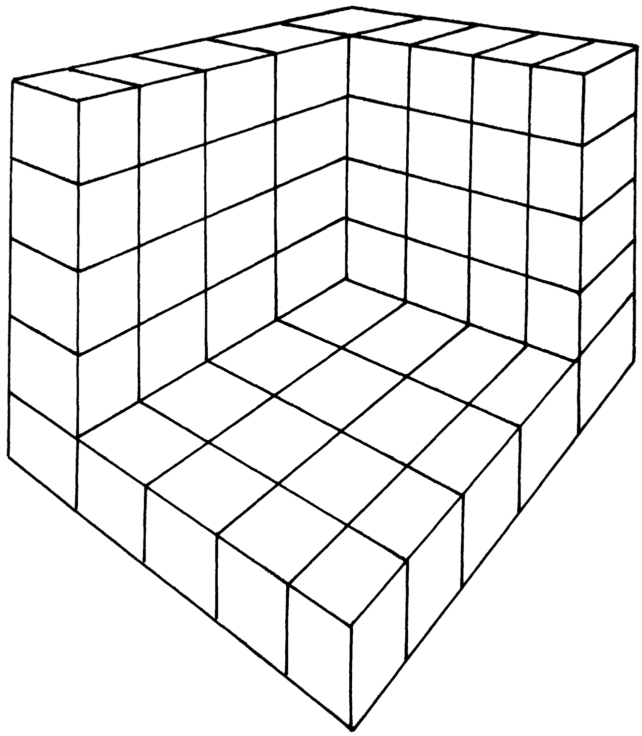


FIG. 6. The fifth hex number.

22. The initial pattern is explained by the facts that if n is odd, $n^4 + 1$ is even, and $17 \times 2^n - 1$ is a multiple of 3. Thereafter it's largely coincidence until $n = 24$, for which $n^4 + 1 = 331777$ is prime, while $17 \times 2^n - 1 = 285212671 = 149 \times 1914179$. See [**17**], [**24**] and sequences 386 and 387 in [**28**].

23. This is also a coincidence, until we reach $n = 18$, for which $21 \times 2^n - 1 = 5505023$ is prime, while

$$7 \times 4^n + 1 = 481036337153 = 166609 \times 2887217.$$

See [**31**], [**23**] and sequences 314 & 315 in [**28**].

24. A sequence introduced by Fritz Göbel. A more convenient recursion for calculation is $(n+1)x_{n+1} = x_n(x_n + n), (n \geqslant 1)$. If you work modulo 43, you'll find that for

n	$=0$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
x_n	$\equiv 1$	2	3	5	10	28	25	37	10	20	15	38	19	42	36	34	2	35	39	31	13	2
n	$= 22$	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	
x_n	$\equiv 6$	26	28	29	4	14	42	5	20	17	4	20	16	29	42	13	42	20	8	23	33	

and $x_{42}(x_{42} + 42) \equiv -10(-10 + 42) = -320$, which is not divisible by 43, so x_{43} is not an integer, although x_n is an integer for $0 \leqslant n \leqslant 42$.

25. Similar calculations, mod 89, using the relation $(n+1)y_{n+1} = y_n(y_n^2 + n)$, show that y_{89} is not an integer. For this, and the previous example, see E15 is [**12**].
26. Since this question was asked, Henry Ibstedt has made extensive calculations, and found the first noninteger term, x_n, in the sequence involving kth powers, to be

k	2	3	4	5	6	7	8	9	10	11
n	43	89	97	214	19	239	37	79	83	239

He also found corresponding results with different initial values. The longest to hold out ($n = 610$) are the cubes ($k = 3$, Example **25**) with $x_0 = 1$, $x_1 = 11$.
27. The first cyclotomic polynomial to display a coefficient other than ± 1 and 0 is

$$\begin{aligned}\Phi_{105}(x) = {} & x^{48} + x^{47} + x^{46} - x^{43} - x^{42} - 2x^{41} - x^{40} - x^{39} + x^{36} + x^{35} + x^{34} \\ & + x^{33} + x^{32} + x^{31} - x^{28} - x^{24} - x^{22} - x^{20} + x^{17} + x^{16} + x^{15} + x^{14} \\ & + x^{13} + x^{12} - x^{9} - x^{8} - 2x^{7} - x^{6} - x^{5} + x^{2} + x + 1\end{aligned}$$

Coefficients can be unboundedly large, but require n to contain a large number of distinct odd prime factors; see [**8**]. More recently, Montgomery & Vaughan [**33**] have shown that if $\Phi_n = \Sigma a(m,n)x^m$ and $L(n) = \ln \max_n |a(m,n)|$ then, for m large, $\frac{m^{1/2}}{(\ln 2m)^{1/4}} \ll L(n) \ll \frac{m^{1/2}}{(\ln m)^{1/4}}$.
28. This game was misremembered by John Conway from John Isbell's game of Beanstalk [**13**]. The Fibonacci pattern is not maintained: only 52 of the first 89 numbers, 81 of the first 144, 126 of the first 233, and 201 of the first 377, are $\mathscr{N}$-positions. The probability argument is fallacious: the probabilities of the status of the two options are *not* independent.
29. True, but why the coincidence?
30 and **31**. The patterns of powers of 2 and of binary digits of $\sqrt{2}$ both continue; see [**11**], [**14**] and sequence 206 in [**28**].
32. A different sequence, number 207 in [**28**], which agrees for $n < 9$, but then continues $28, 41, 60, 88, 129, 189, 277, 406, 595, 872, 1278, \ldots$
33. If $y = x^x$ and $y_n(1)$ denotes the value of $d^n y/dx^n$ at $x = 1$, then

$$\begin{aligned}y_{n+1}(1) = {} & y_n(1) + \binom{n}{1}y_{n-1}(1) - \binom{n}{2}y_{n-2}(1) + 2!\binom{n}{3}y_{n-3}(1) - 3!\binom{n}{4}y_{y-4}(1) \\ & + - + \cdots + (-1)^n(n-1)!.\end{aligned}$$

This was not known to be a multiple of $n+1$ when it was submitted to the Unsolved Problems section of this MONTHLY by Richard Patterson & Gaurar Suri. But in an 87-05-28 letter, Herb Wilf gives a proof, using the generating function for Stirling numbers of the first kind. His proof in fact shows that $n(n-1)$ divides $y_n(1)$ just if $n-1$ divides $(n-2)!$, which it does for $n \geqslant 7$, provided that $n-1$ is not prime.
34. This sequence was investigated by Jim Propp. Except that $a(12) = 55$, the pattern of Fibonacci numbers does not continue:

$n =$ 11	12	13	14	15	16	17	18
$a(n) =$ 35	55	93	149	248	403	670	1082

Since this was written, Wilf [**21**] has linked the generating function with Ramanujan's continued fraction, and he observes that the numbers of propper partitions with k coins in the lowest row are yet another manifestation of the Catalan numbers,

$1, 2, 5, 14, 42, \ldots$ [7]. These partitions are a variant of some considered by Auluck [1]. Auluck's partitions have the pennies contiguous in *every* row, not just the lowest. Their numbers $1, 1, 2, 3, 5, 8, \ldots$ are another good example of the Strong Law.

35. The expansion of the product as a power series, is

$$1 + x^2 + x^3 + x^4 + 2x^5 + 2x^6 + 3x^7 + 3x^8 + 4x^9 + 5x^{10} + 6x^{11} + 7x^{12} + 9x^{13}$$
$$10x^{14} + 12x^{15} + 14x^{16} + 17x^{17} + 19x^{18} + 23x^{19} + 26x^{20} + 30x^{21} + 35x^{22}$$
$$+40x^{23} + 46x^{24} + 52x^{25} + 60x^{26} + 67x^{27} + 77x^{28} + 87x^{29} + \cdots .$$

The sum is the same, until... $+31x^{21} + 35x^{22}$

$$+41x^{23} + 46x^{24} + 54x^{25} + 60x^{26} + 69x^{27} + 78x^{28} + 89x^{29} + \cdots .$$

This was entry 29 in Chapter 5 of Ramanujan's second notebook [2], [3]: but he had crossed it out!

Let me know if I've missed out your favorite example!

REFERENCES

1. F. C. Auluck, On some new types of partitions associated with generalized Ferrers graphs, *Proc. Cambridge Philos. Soc.*, 47 (1951) 679–686; MR 13, 536.
2. Bruce C. Berndt, Ramanujan's Notebooks, Part 1, Springer-Verlag, 1985, p. 130.
3. Bruce C. Berndt and B. M. Wilson, Chapter 5 of Ramanujan's second notebook, in M. I. Knopp (ed.) Analytic Number Theory, Lecture Notes in Math. 899, Springer, 1981, pp. 49–78; MR 83i:10011.
4. W. W. L. Chen, On the error term of the prime number theorem and the difference between the number of primes in the residue classes modulo 4, *J. London Math. Soc.*, (2) 23(1981) 24–40; MR 82g:10058.
5. John Conway and Richard Guy, The Book of Numbers, Scientific American Library, W. H. Freeman, 1988.
6. H. Davenport, The Higher Arithmetic, Hutchinson's University Library, 1952, p. 128.
7. Roger B. Eggleton and Richard K. Guy, Catalan Strikes again! How likely is a function to be convex?, *Math. Mag.* 61(1988) 211–218.
8. P. Erdős and R. C. Vaughan, Bounds for the r-th coefficients of cyclotomic polynomials, *J. London Math. Soc.*, (2) 8(1974) 393–400; MR 50#9835.
9. Martin Gardner, Mathematical games: patterns in primes are a clue to the strong law of small numbers, *Sci. Amer.*, 243 #6(Dec. 1980) 18, 20, 24, 26, 28.
10. Solomon W. Golomb, The evidence for Fortune's conjecture, *Math. Mag.* 54(1981) 209–210.
11. R. L. Graham and H. O. Pollak, Note on a linear recurrence related to $\sqrt{2}$, *Math. Mag.*, 43(1970) 143–145; MR42 #180.
12. Richard K. Guy, Unsolved Problems in Number Theory, Springer, 1981.
13. Richard K. Guy, John Isbell's game of Beanstalk and John Conway's game of Beans-Don't-Talk, *Math. Mag.*, 59(1986) 259–269.
14. F. K. Huang and S. Lin, An analysis of Ford & Johnson's sorting algorithm, *Proc. 3rd Annual Princeton Conf. on Info. Systems and Sci.*
15. Wilfrid Keller, Factors of Fermat numbers and large primes of the form $k.2^{n+1}$, *Math. Comput.*, 41(1983) 661–673.
16. R. B. Killgrove and K. E. Ralston, On a conjecture concerning the primes, *Math. Tables Aids Comput.*, 13(1959) 121–122; MR 21#4943.
17. M. Lal, Primes of the form $n^4 + 1$, *Math. Comput.*, 21(1967) 245–247.
18. J. E. Littlewood, Sur le distribution des nombres premiers, *C. R. hebd. Séanc. Acad. Sci., Paris*, 158(1914) 1868–1872.
19. Calvin T. Long, Strike it out–add it up, *Math. Mag.*, 66(1982) 273–277. See also this MONTHLY 73(1966) 846–851.
20. Alfred Moessner, Eine Bemerkung über die Potenzen der natürlichen Zahlen, *S.-B. Math.-Nat. Kl. Bayer. Akad. Wiss.*, 1951, 29(1952); MR 14-353b.

21. Andrew M. Odlyzko and Herbert S. Wilf, n coins in a fountain (to appear in *Amer. Math. Monthly.* (Nov. 1988)).
22. Ivan Paasche, Eine Verallgemeinerung des Moessnerschen Satzes, *Compositio Math.*, 12(1956) 263–270; MR 17, 836g.
23. Hans Riesel, Lucasian criteria for the primality of $N = h.2^n - 1$, *Math. Comput.*, 23(1969) 869–875.
24. Raphael M. Robinson, A report on primes of the form $k \cdot 2^n + 1$ and on factors of Fermat numbers, *Proc. Amer. Math. Soc.*, 9(1958) 673–681.
25. P. Seeling, Verwandlung der irrationalen Grösse $\sqrt[n]{\ }$ in einen Kettenbruch, *Archiv. Math. Phys.*, 46(1866) 80–120 (esp. p. 116).
26. J. L. Selfridge, Solution to problem 4995, this MONTHLY, 70(1963) 101.
27. W. Sierpiński, Sur un problème concernant les nombres $k \cdot 2^n + 1$, *Elem. Math.*, 15(1960) 73–74; MR 22 #7983; corrigendum, *ibid.* 17(1962) 85.
28. N. J. A. Sloane, A Handbook of Integer Sequences, Academic Press, 1973.
29. Harold M. Stark, An explanation of some exotic continued fractions found by Brillhart, Computers in Number Theory, Atlas Sympos. No. 2, Oxford, 1969, pp. 21–35, Academic Press, London, 1971.
30. Peter Taylor and Doug Dillon, Problem 3, *Queen's Math. Communicator*, Dept. of Math. and Statist., Queen's University, Kingston, Ont., Oct. 1985, p. 16.
31. H. C. Williams and C. R. Zarnke, A report on prime numbers of the forms $M = (6a + 1)2^{2m-1} - 1$ and $M' = (6a - 1)2^{2m} - 1$, *Math. Comput.*, 22(1968) 420–422.
32. Jeff Young and Duncan A. Buell, The twentieth Fermat number is composite, *Math. Comput.* 50(1988) 261–263.
33. H. L. Montgomery and R. C. Vaughan, The order of magnitude of the mth coefficients of cyclotomic polynomials, *Glasgow Math. J.* 27(1985) 143–159; MR 87e: 11026.

A BIRTHDAY GREETING

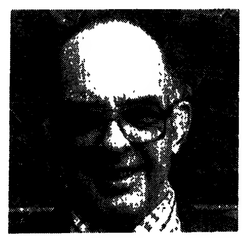
The readers and editors of the MONTHLY send their warm best wishes to Professor I. J. Schoenberg, who has contributed so much to mathematics in general, and to the MONTHLY, in particular, on the occasion of his 85th birthday.

The American Romanian Academy has recognized the occasion by conferring on him an award, "In recognition of his distinguished contributions to the advancement of scholarship in the sciences, in the spirit of free exchanges of values and ideas." His publisher, Birkhäuser, has recognized the occasion by publishing two volumes of Dr. Schoenberg's mathematical writings, edited by Carl de Boor.

Every Smooth Map of Euclidean Space into Itself Is an Expansion Followed by a Contraction

JOHN MITCHELL, *Prime Computer*
LEE A. RUBEL, *University of Illinois*

JOHN MITCHELL received his Ph.D. in mathematics from the State University of New York at Stony Brook in 1982. He was a Hedrick Assistant Professor at U.C.L.A. from 1982 to 1985, and he was a visiting assistant Professor at the University of Illinois, Urbana-Champaign from 1985–7. His research interests are in differential geometry, specifically, sub-Riemannian geometry. He is now at Prime Computer, Framingham, MA.



LEE A. RUBEL is Professor of Mathematics at the University of Illinois at Urbana-Champaign, where he has been for thirty years, except for scholarly leaves. He got his Ph.D. at the University of Wisconsin (Madison) under R. C. Buck's direction in 1954, then spent two years at Cornell University and two years at the Institute for Advanced Study before coming to Illinois. His earlier work was mostly in Complex Variables, but his recent and current work is mostly in Algebraic Differential Equations. He has published about 160 research articles and two books, and has lectured at many universities around the world. He has been an editor of several journals, and is currently an Associate Editor of the MONTHLY.

Dedicated to the memory of Irving Reiner

Abstract. Let $n \geqslant 1$ and F: $\mathbb{R}^n \to \mathbb{R}^n$ be any C^1 map (into). Then F may be written $F = C \circ E$, where E is an expansion (i.e. $\|E(x) - E(x')\| \geqslant \|x - x'\|$ for all $x, x' \in \mathbb{R}^n$) and C is a contraction (i.e. $\|C(x) - C(x')\| \leqslant \|x - x'\|$ for all $x, x' \in \mathbb{R}^n$.) For $n \geqslant 2$, there is a diffeomorphism F: $\mathbb{R}^n \xrightarrow{onto} \mathbb{R}^n$ that cannot be written $F = E \circ C$. Some related results are proved. For example, it is shown that the "typical" continuous map F: $\mathbb{R}^1 \to \mathbb{R}^1$ can neither be written $F = C \circ E$ nor $F = E \circ C$.

We prove some theorems about factoring maps of Euclidean n-space into itself as a composition of an expansion and a contraction. Our results can be given a cosmological flavor (if we believe the universe is $\mathbb{R}^3$), by saying that the present state of the universe could have been achieved as an expansion followed by a contraction, but possibly not as a contraction followed by an expansion. The proofs of our Theorems 5 and 6 show why we require that the map F to be factored satisfy some differentiability requirements, for if F is everywhere badly nondifferentiable, no such factorization is possible (see Remark 3).

THEOREM 1. *Let F be a C^1 mapping of $\mathbb{R}^n$ into $\mathbb{R}^n$. Then F is a composition of a C^1 expansion E followed by a C^1 contraction, $F = C \circ E$.*

Proof. Writing $d(p, q) = \|p - q\|$, we must show that there is a C^1 expansion E of $\mathbb{R}^n$ for which $F \circ E^{-1}$ is a C^1 contraction. (It is easy to verify that if E is an expansion, then it must be a homeomorphism.) That is, for all $p, q \in \mathbb{R}^n$,

$d(F(E^{-1}(p)), F(E^{-1}(q))) \leqslant d(p,q)$, which is equivalent to: for all $P, Q \in \mathbf{R}^n$,

$$d(F(P), F(Q)) \leqslant d(E(P), E(Q)).$$

Set $P = E^{-1}(p)$, $Q = E^{-1}(q)$. So we need to find an expansion E that expands more than F does. For this purpose, it suffices to find an expansion E which is *infinitesimally* more expanding than F is, by which we mean that the differential E_* of E increases the length of each tangent vector to $\mathbf{R}^n$ more than the differential of F does. For if we have found such an E, then E increases the length of any curve more than F does. Then, if we denote by γ the inverse image under E of the straight line segment joining $E(P)$ to $E(Q)$, then

$$d(F(P), F(Q)) \leqslant \text{length}\,(F(\gamma)) \leqslant \text{length}\,(E(\gamma)) = d(E(P), E(Q)).$$

This first part of the proof holds quite generally—for example on any Riemannian manifold. It remains to show that there is, for any $n \geqslant 1$, a C^1 expansion E which is infinitesimally at least as expanding as the given C^1 mapping F.

We may choose E to have the form

$$E(p) = \lambda(r) \cdot p,$$

where $r = \|p\|$ is the distance to the origin, and $\lambda(r)$ is any smooth function defined for $r \geqslant 0$ and satisfying

(i) $\lambda(r)$ is increasing

(ii) $\lambda(r) \geqslant \max\{\|F_*(v_x)\|: \|x\| = r,\ v_x$ is a unit tangent vector to x in $\mathbf{R}^n\}$, and

(iii) $\lambda(r) \geqslant 1$.

That such functions λ exist is obvious. That the associated mapping E is expanding (infinitesimally and hence globally also since E will be one : one) follows from (iii), just as does the fact that E expands more than F does, follows from (ii), once we see that (i) implies that

$$\|E_*(v_x)\| \geqslant \lambda(\|x\|) \cdot \|v_x\|$$

for any tangent vector v_x at x. To prove this last fact, we may assume that v_x points outward, i.e. that $v_x \cdot x \geqslant 0$, since E_* is linear. (Otherwise replace v_x by $-v_x$.) Now if $y \in \mathbf{R}^n$ satisfies $\|y\| \geqslant \|x\|$, then E may be thought of as acting on the pair x, y by first multiplying by $\lambda(\|x\|)$ so that

$$x \mapsto \bar{x} = \lambda(\|x\|) \cdot x \quad \text{and} \quad y \mapsto \bar{y} = \lambda(\|x\|) \cdot y,$$

and then pushing $\bar{y}$ out even farther, since $\lambda(\|y\|) \geqslant \lambda(\|x\|)$. The homothety by $\lambda(\|x\|)$ increases distances, since $\lambda \geqslant 1$, and pushing $\bar{y}$ out farther increases its distance from $\bar{x}$ still more, since $\|\bar{y}\| \geqslant \|x\|$, as a simple geometrical argument in the plane through 0, x, and y shows. So $d(x, y)$ is increased at least by a factor $\lambda(\|x\|)$ if $\|y\| \geqslant \|x\|$. For y infinitesimally close to x, this says that $\|E_*(v_x)\| \geqslant \lambda(\|x\|) \cdot \|v_x\|$ for any tangent vector at x, and we are done.

Notice that we may choose $\lambda(r)$ to be a constant for small r, which will guarantee that E is differentiable at the origin. Also, $\lambda(r)$, and hence E, may be chosen to be C^∞, so that the contraction C will have as many derivatives as F does.

Remark 1. The same method of proof will work on any complete simply-connected open Riemannian manifold of nonpositive sectional curvature. Notice however that some open Riemannian simply-connected manifolds do not admit *any* expanding map that strictly expands some distance—for example a manifold with finite volume cannot admit such an expansion.

THEOREM 2. *There exists, if $n \geqslant 2$, a C^∞ diffeomorphism F of $\mathbb{R}^n$ onto $\mathbb{R}^n$ which may not be written as $F = E \circ C$, where C is a contraction of $\mathbb{R}^n$ and E is an expansion.*

Proof. Define, for each $t \geqslant 1$, H_t: $S^{n-1} \to S^{n-1}$ to be the dilation by the factor t on $\mathbb{R}^{n-1}$, transferred to S^{n-1} by stereographic projection. Thus, points get pushed away from the South pole, towards the North pole. Now define F: $\mathbb{R}^n \to \mathbb{R}^n$ to act as the application of H_{t+1} to the sphere $S^{n-1}(t)$ of radius t centered at the origin in $\mathbb{R}^n$. (Thus, we think of $\mathbb{R}^n$ as an onion, with the t-th layer being $S^{n-1}(t)$.) If necessary, smooth F near the origin.

If we were to have $F = E \circ C$, then we would have $E^{-1} = C \circ F^{-1}$, so that the contraction E^{-1} would be more contracting than F^{-1} is. Consider now what E^{-1} does to $S^{n-1}(t)$, the sphere of radius t centered at 0, when t is large. Since F^{-1} contracts the complement of some small neighborhood of the North pole on $S^{n-1}(t)$ to a small neighborhood of the South pole, and since E^{-1} must contract at least as much as F, the $(n-1)$-volume of E^{-1} of this complement must be at most equal to the $(n-1)$-volume of F^{-1} of this complement, which is very small. Now since E^{-1} is a contraction, the $(n-1)$-volume of E^{-1} of that small neighborhood of the North pole is also very small. Consequently, the $(n-1)$-volume of $E^{-1}(S^{n-1}(t))$ is very small if t is large. By the isoperimetric inequality, we conclude that the n-volume of $E^{-1}(B^n(t))$ is very small if t is large, where

$$B^n(t) = \{x \in \mathbb{R}^n\colon \|x\| \leqslant t\}.$$

Thus

$$\lim_{t\to\infty} \operatorname{vol}\big(E^{-1}(B^n(t))\big) = 0. \qquad (*)$$

But the $B^n(t)$'s increase as sets as t increases, so the same is true of $E^{-1}(B^n(t))$. Consequently, vol$(E^{-1}(B^n(t)))$ must increase with t. However, vol $E^{-1}(B^n(t)) > 0$, for all t, since E^{-1} is one : one and continuous, and thus E^{-1} is an open map. This contradicts $(*)$ and shows that no such E exists.

Remark 2. The above proof actually shows that F cannot be written as $F = e \circ c$, where c reduces the $(n-1)$-volumes of hypersurfaces and e locally expands the $(n-1)$-volumes of hypersurfaces. (Here we think of the domain of e as the range of c.) Notice that, by the isoperimetric inequality, such mappings c and e would have to reduce, respectively expand, n-volumes locally. This raises the question of whether we can write such an F as $F = \varepsilon \circ \gamma$, where γ reduces n-volumes, and ε locally expands them.

THEOREM 3. *There exists a C^∞ map F: $\mathbb{R}^1 \to \mathbb{R}^1$ that cannot be written $F = E \circ C$ where C is a contraction and E is an expansion.*

Proof. Let $p_i = i$, $q_i = i + 1/i$, $i = 2, 3, 4, \ldots$. Let F be a C^∞ map on $\mathbb{R}^1$ such that $F(p_i) = 0$ and $F(q_i) = 1$ for all i. (Actually, we may take F real-analytic.) If we could write $F = E \circ C$, then we would have $E(C(p_i)) = 0$, $E(C(q_i)) = 1$ for all i. Since E would be a homeomorphism, we would have $C(p_i) = E^{-1}(0)$, $C(q_i) = E^{-1}(1)$. But $d(E^{-1}(0), E^{-1}(1)) \leqslant d(p_i, q_i) \to 0$, and we thus have a contradiction.

THEOREM 4. *Let F: $\mathbb{R}^1 \to \mathbb{R}^1$ be a one : one C^1 map. Then there exist a C^1 contraction C and a C^1 expansion E such that $F = E \circ C$.*

Proof. It clearly suffices to find $g, h \in C^1(\mathbb{R}^1)$ such that $g'(x) \geqslant 1$ for all x, $|h'(x)| \leqslant 1$ for all x and $F = g \circ h$. To do this, let $h'(x) = F'(x)$ where $|F'(x)| \leqslant 1$, and define $h'(x) = \operatorname{sgn} F'(x)$ where $|F'(x)| \geqslant 1$, and let $h(0) = 0$. This defines $h(x)$. If $|F'(x)| \leqslant 1$, define $g'(h(x)) = 1$. If $|F'(x)| \geqslant 1$, let $g'(h(x)) = |F'(x)|$. Let $g(0) = F(0)$. This defines $g(x)$. Now $F = g \circ h$ since $F(0) = g(h(0))$ and $F'(x) = g'(h(x))h'(x)$ for all x. Q.E.D.

THEOREM 5. *There exists a continuous map F: $\mathbb{R} \to \mathbb{R}$ that cannot be written $F = C \circ E$, where C is a contraction and E is an expansion.*

Proof. Let F be any nowhere differentiable continuous function, and suppose $F = C \circ E$. Since E is an expansion, it is a monotonic function on $\mathbb{R}$, and consequently its derivative $E'(x)$ exists and is finite on a set A whose complement has zero measure. Since C is a Lipschitz function, $C'(x)$ exists and is finite off a set B of zero measure. As x runs over A, $E(x)$ certainly runs over a set of positive measure. Consequently $\frac{d}{dx}(C \circ E)(x) = C'(E(x))E'(x)$ exists and is finite for some $x \in \mathbb{R}$, which is impossible if $F = C \circ E$.

THEOREM 6. *There exists a continuous map F: $\mathbb{R} \to \mathbb{R}$ that cannot be written $F = E \circ C$, where E is an expansion and C is a contraction.*

Proof. Let F be the Weierstrass function:

$$F(x) = \sum_{n=1}^{\infty} a^n \cos(\pi b^n x);\ 0 < a < 1,\ ab > \frac{3}{2}\frac{\pi}{1 - a},$$

where b is an odd integer (see [**1**, pp. 401–404]). We will actually prove that we may not write $F = H \circ C$ where H is a homeomorphism, and C is a contraction. For there must exist an $x \in \mathbb{R}$ such that $H'(C(x))$ exists and is finite, because H is monotone and the range of C must contain a nontrivial interval. For this x (see [**1**, pp. 404]—note that one actually has there $N_n \geqslant \delta > 0$ and not merely $N_n > 0$, for all n) there exists a sequence $x_n \to x$ such that $|f(x_n) - f(x)|/|x_n - x| \to \infty$. Clearly, $C(x_n) = C(x)$ for at most finitely many n, which exceptions we discard. Then

$$\left|\frac{f(x_n) - f(x)}{x_n - x}\right| = \left|\frac{H(C(x_n)) - H(C(x))}{C(x_n) - C(x)}\right| \cdot \left|\frac{C(x_n) - C(x)}{x_n - x}\right|.$$

As $n \to \infty$, the left-hand side approaches ∞, while the first term on the right-hand side approaches $|H'(C(x))| < \infty$. But the last term is bounded by 1, which is clearly impossible.

Remark 3. By a result of Jarnik [**2**], every continuous function F: $\mathbb{R} \to \mathbb{R}$, with the exception of a set of functions of the first category in $C(\mathbb{R})$, has *every* extended real number as a derived number at every point. So the above proofs show that the 'typical' continuous map F: $\mathbb{R} \to \mathbb{R}$ *cannot* be written either as $F = E \circ C$ or $F = C \circ E$.

REFERENCES

1. E. W. Hobson, The Theory of Functions of a Real Variable and the Theory of Fourier's Series, Vol II, Dover Publications, Inc., New York, 1957.
2. V. Jarnik, Über die Differenzierbarkeit stetiger Funktionen, *Fund. Math.*, 21 (1933) 48–58.

The William Lowell Putnam Mathematical Competition

LEONARD F. KLOSINSKI, Santa Clara University

G. L. ALEXANDERSON, Santa Clara University

LOREN C. LARSON, St. Olaf College

The following results of the forty-eighth William Lowell Putnam Mathematical Competition, held on December 5, 1987, have been determined in accordance with the governing regulations. This annual contest is supported by the William Lowell Putnam Prize Fund for the Promotion of Scholarship, left by Mrs. Putnam in memory of her husband, and is held under the auspices of the Mathematical Association of America.

The first prize, $5,000, was awarded to the Department of Mathematics of Harvard University. The members of the winning team were: David J. Moews, Bjorn M. Poonen, and Michael Reid; each was awarded a prize of $250.

The second prize, $2,500, was awarded to the Department of Mathematics of Princeton University. The members of the winning team were: Daniel J. Bernstein, David J. Grabiner, and Matthew D. Mullin; each was awarded a prize of $200.

The third prize, $1,500, was awarded to the Department of Mathematics of Carnegie-Mellon University. The members of the winning team were: Petros I. Hadjicostas, Joseph G. Keane, and Karl M. Westerberg; each was awarded a prize of $150.

The fourth prize, $1,000, was awarded to the Department of Mathematics of the University of California, Berkeley. The members of the winning team were: David P. Moulton, Jonathan E. Shapiro, and Christopher S. Welty; each was awarded a prize of $100.

The fifth prize, $500, was awarded to the Department of Mathematics of the Massachusetts Institute of Technology. The members of the winning team were: David T. Blackston, James P. Ferry, and Waldemar P. Horwat; each was awarded a prize of $50.

The six highest-ranking individual contestants, in alphabetical order, were David J. Grabiner, Princeton University; David J. Moews, Harvard University; Bjorn M. Poonen, Harvard University; Michael Reid, Harvard University; Constantin S. Teleman, Harvard University; and John S. Tillinghast, University of California, Davis. Each of these students was designated a Putnam Fellow by the Mathematical Association of America and awarded a prize of $500 by the Putnam Prize Fund.

The next five highest-ranking individuals, in alphabetical order, were Daniel J. Bernstein, Princeton University; Constantine N. Costes, Harvard University; Jeremy A. Kahn, Harvard University; Rav Kumar Ramakrishna, Cornell University; and Japheth Wood, Washington University, St. Louis. Each was awarded a prize of $250.

The following teams, named in alphabetical order, received honorable mention: Rice University, with team members Charles R. Ferenbaugh, Thomas M. Hyer, and John W. McIntosh; Stanford University, with team members Thomas H. Chung, John A. Overdeck, and Joshua R. Zucker; the University of Toronto, with team members Gary F. Baumgartner, Edward I. Doolittle, and Jeffrey S. Rosenthal;

Washington University, St. Louis, with team members Daniel N. Ropp, Peter Shawhan, and Japheth Wood; and Yale University, with team members Kamal F. Khuri-Makdisi, Robert S. Manning, and William M. Nelson.

Honorable mention was achieved by the following thirty-eight individuals named in alphabetical order: Thomas R. Amoth, Oregon State University; Leland F. Brown, California Institute of Technology; Emory F. Bunn, Princeton University; Yeow Meng Chee, University of Waterloo; Timothy Y. Chow, Princeton University; William P. Cross, California Institute of Technology; Galia D. Dafni, Pennsylvania State University; Douglas R. Davidson, Princeton University; Frank M. D'Ippolito, University of Waterloo; Edward J. Doolittle, University of Toronto; Serge Elnitsky, Carleton University; Chenteh Kenneth Fan, Harvard University; Petros Hadjicostas, Carnegie-Mellon University; Thomas R. Hagedorn, Princeton University; Thomas S. Harke, University of Alberta; Waldemar P. Horwat, Massachusetts Institute of Technology; William C. Jockusch, Carleton College; Alex T. Kachura, University of Waterloo; Joseph G. Keane, Carnegie-Mellon University; Kamal F. Khuri-Makdisi, Yale University; Jordan Lampe, University of California, Berkeley; Daniel D. Lee, Harvard University; John W. McIntosh, Rice University; Robert S. Manning, Yale University; David P. Moulton, University of California, Berkeley; Matthew D. Mullin, Princeton University; Du Nguyen, Ottawa University; David L. Petry, University of Oregon; Daniel N. Ropp, Washington University, St. Louis; David B. Secrest, University of Illinois, Urbana-Champaign; Robert G. Southworth, California Institute of Technology; Glenn P. Tesler, California Institute of Technology; Martin Trudeau, Université de Montréal; Christopher S. Welty, University of California, Berkeley; Karl M. Westerberg, Carnegie-Mellon University; Glen T. Whitney, Harvard University; Russil Wvong, University of British Columbia; and Joshua R. Zucker, Stanford University.

The other individuals who achieved ranks among the top 101, in alphabetical order of their schools, were: Amherst College, Peter H. Anspach; Bethel College, Jonathan P. McCammond; University of British Columbia, Wayne J. Broughton; Brown University, Kevin S. McFarland, David J. Morin; California Institute of Technology, Jared C. Bronski, Philip W. Nabours; University of California, Berkeley, Jonathan E. Shapiro; University of California, Los Angeles, Joseph M. Rojas; University of California, San Diego, David L. Ruhm; Carleton University, Michael J. Bradley, Stephen A. Smith; Case Western Reserve University, Patrick T. Headley, William E. Kirby; University of Chicago, Linda E. Green, Robert P. Stingley, Andrew S. Yeh; Dalhousie University, Daniel J. Peters; Harvard University, Michael J. Callahan, David Cook, Michael P. Mitzenmacher; University of Hawaii, Jor-Kuen E. Lo; University of Houston (Clear Lake), John Ken Burton, Jr.; University of Illinois, Urbana-Champaign, James M. Grochocinski; University of Maryland, Catonsville, Michael J. Johnson; Massachusetts Institute of Technology, David T. Blackston, Claudio C. Chamon, James P. Ferry, Mark Kantrowitz, James R. Rauen; University of Michigan, Ann Arbor, William F. Doran IV, Matthew A. Klimesh, Paul Kominsky; Oberlin College, David B. Carlton; University of Pennsylvania, Michael Albert; Princeton University, Rahul V. Pandharipande; University of Regina, Simon H. Lee; Rice University, Charles R. Ferenbaugh, Thomas M. Hyer; Siena College, Gregory S. Spradlin; University of South Carolina, Roger B. Milne; Stanford University, John C. Loftin, John A. Overdeck; University of Texas, Austin, Jared L. Levy; University of Toronto, Gary F. Baumgartner, Jeffrey S.

Rosenthal; Virginia Polytechnic Institute and State University, Patrick R. Brown; Washington and Lee University, John D. Boller; Washington University, St. Louis, David S. Shobe; University of Waterloo, Marc A. Chamberland, Bryan K. Feir, Giuseppe Russo.

There were 2170 individual contestants from 359 colleges and universities in Canada and the United States in the competition of December 5, 1987. Teams were entered by 277 institutions.

The Questions Committee for the forty-eighth competition consisted of Harold M. Stark (Chairman), Gerald A. Heuer, and Abraham P. Hillman; they composed the problems listed below and were most prominent among those suggesting solutions.

PROBLEMS

Problem A-1

Curves A, B, C, and D, are defined in the plane as follows:

$$A = \left\{(x, y) : x^2 - y^2 = \frac{x}{x^2 + y^2}\right\},$$

$$B = \left\{(x, y) : 2xy + \frac{y}{x^2 + y^2} = 3\right\},$$

$$C = \left\{(x, y) : x^3 - 3xy^2 + 3y = 1\right\},$$

$$D = \left\{(x, y) : 3x^2y - 3x - y^3 = 0\right\}.$$

Prove that $A \cap B = C \cap D$.

Problem A-2

The sequence of digits

$$1\,2\,3\,4\,5\,6\,7\,8\,9\,1\,0\,1\,1\,1\,2\,1\,3\,1\,4\,1\,5\,1\,6\,1\,7\,1\,8\,1\,9\,2\,0\,2\,1\ldots$$

is obtained by writing the positive integers in order. If the 10^nth digit in this sequence occurs in the part of the sequence in which the m-digit numbers are placed, define $f(n)$ to be m. For example, $f(2) = 2$ because the 100th digit enters the sequence in the placement of the two-digit integer 55. Find, with proof, $f(1987)$.

Problem A-3

For all real x, the real-valued function $y = f(x)$ satisfies

$$y'' - 2y' + y = 2e^x.$$

(a) If $f(x) > 0$ for all real x, must $f'(x) > 0$ for all real x? Explain.
(b) If $f'(x) > 0$ for all real x, must $f(x) > 0$ for all real x? Explain.

Problem A-4

Let P be a polynomial, with real coefficients, in three variables and F be a function of two variables such that

$$P(ux, uy, uz) = u^2F(y - x, z - x) \quad \text{for all real } x, y, z, u,$$

and such that $P(1,0,0) = 4$, $P(0,1,0) = 5$, and $P(0,0,1) = 6$. Also let A, B, C be complex numbers with $P(A, B, C) = 0$ and $|B - A| = 10$. Find $|C - A|$.

Problem A-5
Let

$$\vec{G}(x, y) = \left(\frac{-y}{x^2 + 4y^2}, \frac{x}{x^2 + 4y^2}, 0\right).$$

Prove or disprove that there is a vector-valued function

$$\vec{F}(x, y, z) = (M(x, y, z), N(x, y, z), P(x, y, z))$$

with the following properties:
(i) M, N, P have continuous partial derivatives for all $(x, y, z) \neq (0,0,0)$;
(ii) Curl $\vec{F} = \vec{0}$ for all $(x, y, z) \neq (0,0,0)$;
(iii) $\vec{F}(x, y, 0) = \vec{G}(x, y)$.

Problem A-6
For each positive integer n, let $a(n)$ be the number of zeros in the base 3 representation of n. For which positive real numbers x does the series

$$\sum_{n=1}^{\infty} \frac{x^{a(n)}}{n^3}$$

converge?

Problem B-1
Evaluate

$$\int_2^4 \frac{\sqrt{\ln(9 - x)}\, dx}{\sqrt{\ln(9 - x)} + \sqrt{\ln(x + 3)}}.$$

Problem B-2
Let r, s, and t be integers with $0 \leqslant r$, $0 \leqslant s$, and $r + s \leqslant t$. Prove that

$$\frac{\binom{s}{0}}{\binom{t}{r}} + \frac{\binom{s}{1}}{\binom{t}{r+1}} + \frac{\binom{s}{2}}{\binom{t}{r+2}} + \cdots + \frac{\binom{s}{s}}{\binom{t}{r+s}} = \frac{t+1}{(t+1-s)\binom{t-s}{r}}.$$

(Note: $\binom{n}{k}$ denotes the binomial coefficient $n(n-1)\cdots(n+1-k)/k(k-1)\cdots 3 \cdot 2 \cdot 1$.)

Problem B-3
Let F be a field in which $1 + 1 \neq 0$. Show that the set of solutions to the equation $x^2 + y^2 = 1$ with x and y in F is given by $(x, y) = (1,0)$ and

$$(x, y) = \left(\frac{r^2 - 1}{r^2 + 1}, \frac{2r}{r^2 + 1}\right),$$

where r runs through the elements of F such that $r^2 \neq -1$.

Problem B-4
Let $(x_1, y_1) = (0.8, 0.6)$ and let $x_{n+1} = x_n \cos y_n - y_n \sin y_n$ and $y_{n+1} = x_n \sin y_n + y_n \cos y_n$ for $n = 1, 2, 3, \ldots$. For each of $\lim_{n\to\infty} x_n$ and $\lim_{n\to\infty} y_n$, prove that the limit exists and find it or prove that the limit does not exist.

Problem B-5

Let O_n be the n-dimensional zero vector $(0, 0, \ldots, 0)$. Let M be a $2n \times n$ matrix of complex numbers such that whenever $(z_1, z_2, \ldots, z_{2n})M = O_n$, with complex z_i, not all zero, then at least one of the z_i is not real. Prove that for arbitrary real numbers $r_1, r_2, \ldots, r_{2n}$, there are complex numbers $w_1, w_2, \ldots, w_n$ such that

$$\operatorname{Re}\left[M\begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}\right] = \begin{pmatrix} r_1 \\ \vdots \\ r_{2n} \end{pmatrix}$$

(Note: If C is a matrix of complex numbers, $\operatorname{Re}(C)$ is the matrix whose entries are the real parts of entries of C.)

Problem B-6

Let F be the field of p^2 elements where p is an odd prime. Suppose S is a set of $(p^2 - 1)/2$ distinct nonzero elements of F with the property that for each $a \neq 0$ in F, exactly one of a and $-a$ is in S. Let N be the number of elements in the intersection $S \cap \{2a\colon a \in S\}$. Prove that N is even.

SOLUTIONS

In the 12-tuples $(n_{10}, n_9, \ldots, n_0, n_{-1})$ following each problem number below, n_i for $10 \geqq i \geqq 0$ is the number of students among the top 204 contestants achieving i points for the problem and n_{-1} is the number of those not submitting solutions.

A-1. (72, 24, 22, 0, 0, 0, 0, 0, 5, 0, 45, 36)

Solution 1. First note that $(0,0)$ doesn't belong to either $A \cap B$ or $C \cap D$, so in what follows suppose that $(x, y) \neq (0,0)$.

Let $Eq(i)$, $i = 1,2,3,4$, denote the equation that defines the set A, B, C, D respectively. Also, let $f(x, y)Eq(i)$ denote the equation obtained by multiplying each side of $Eq(i)$ by $f(x, y)$.

The matrix product

$$\begin{pmatrix} x & -y \\ y & x \end{pmatrix}\begin{pmatrix} Eq(1) \\ Eq(2) \end{pmatrix} = \begin{pmatrix} Eq(3) \\ Eq(4) \end{pmatrix}$$

shows that $A \cap B \subseteq C \cap D$, and

$$\begin{pmatrix} Eq(1) \\ Eq(2) \end{pmatrix} = \begin{pmatrix} x & -y \\ y & x \end{pmatrix}^{-1}\begin{pmatrix} Eq(3) \\ Eq(4) \end{pmatrix}$$

$$= \begin{pmatrix} \dfrac{x}{x^2+y^2} & \dfrac{y}{x^2+y^2} \\ \dfrac{-y}{x^2+y^2} & \dfrac{x}{x^2+y^2} \end{pmatrix}\begin{pmatrix} Eq(3) \\ Eq(4) \end{pmatrix}$$

shows that $C \cap D \subseteq A \cap B$.

These two inclusions show that $A \cap B = C \cap D$.

Solution 2. Let $z = x + iy$. Then

$$\begin{aligned}(x, y) \in A \cap B \quad &\text{iff } z^2 = 3i + 1/z\\ &\text{iff } z^3 = 3iz + 1\\ &\text{iff } (x, y) \in C \cap D.\end{aligned}$$

A-2. (117, 28, 22, 2, 0, 0, 0, 3, 10, 5, 10, 7)

The r-digit numbers run from 10^{r-1} to $10^r - 1$, so there are $10^r - 10^{r-1}$ of them. Thus, the total number of digits in numbers with at most r digits is $g(r) = \Sigma_{k=1}^{r} k(10^k - 10^{k-1}) = -1 + \Sigma_{k=1}^{r-1}(k - (k+1))10^k + r10^r = -\Sigma_{k=0}^{r-1}10^k + r10^r = r10^r - (10^r - 1)/9$ for $r \geqslant 1$. But $0 < (10^r - 1)/9 < 10^r$, so $(r-1)10^r < g(r) < r10^r$. Thus, $g(1983) < 1983 \cdot 10^{1983} < 10^4 \cdot 10^{1983} = 10^{1987}$, and $g(1984) > 1983 \cdot 10^{1984} > 10^3 \cdot 10^{1984} = 10^{1987}$. It follows that $f(1987) = 1984$.

A-3. (119, 32, 1, 0, 1, 4, 2, 0, 0, 5, 11, 29)

The general solution to the differential equation is $f(x) = (x^2 + bx + c)e^x$ with b and c real. For such a function $f'(x) = (x^2 + (b+2)x + (b+c))e^x$. Clearly $f(x) > 0$ for all x if and only if $D = b^2 - 4ac < 0$, and $f'(x) > 0$ for all x if and only if $D' = (b+2)^2 - 4(b+c) = b^2 + 4b + 4 - 4b - 4c = D + 4 < 0$. The answer to (a) is "no" because $D < 0$ does not imply $D' < 0$. (For example, take $b = c = 1$; then $f(x) > 0$ for all x but $f'(-1) = 0$.) The answer to (b) is "yes" because $D' < 0$ implies $D < 0$.

A-4. (14, 8, 6, 4, 0, 0, 0, 1, 28, 9, 25, 109)

Letting $u = 1$ and $x = 0$, we have $F(y, z) = P(0, y, z)$ is a polynomial. $F(uy, uz) = P(0, uy, uz) = u^2F(y, z)$, so F is homogeneous of degree 2. Now $P(x, y, z) = F(y - x, z - x)$ implies that

$$P(x, y, z) = a(y - x)^2 + b(y - x)(z - x) + c(z - x)^2$$

with a, b, c real. Then $4 = P(1,0,0) = a + b + c$, $5 = P(0,1,0) = a$, and $6 = P(0,0,1) = c$. It follows that $4 = a + b + c = 5 + b + 6$, and so $b = -7$.

For the complex numbers A, B, C of the hypothesis, we have $5(B - A)^2 - 7(B - A)(C - A) + 6(C - A)^2 = 0$. Let $m = (C - A)/(B - A)$. Then $5 - 7m + 6m^2 = 0$. The roots of $6m^2 - 7m + 5 = 0$ are complex, so $|m| = \sqrt{5/6}$. Hence, $|C - A| = \sqrt{5/6}\,|B - A| = (5/3)\sqrt{30}$.

A-5. (8, 0, 2, 0, 0, 0, 0, 0, 1, 1, 57, 135)

Note that Curl $\vec{G} = \vec{0}$ unless (x, y, z) is on the z-axis. If $\vec{F}$ exists, then by Stokes' Theorem,

$$\int_C \vec{G}(x, y) \cdot d\vec{r} = \int_C \vec{F}(x, y, z) \cdot d\vec{r} = \iint_S \left(\text{Curl } \vec{F}\right) \cdot \vec{n}\, dS = 0,$$

where C is the ellipse $x^2 + 4y^2 = 1$, $z = 0$, in the xy-plane and S is the part of the ellipsoid $x^2 + 4y^2 + z^2 = 1$ with $z \geqslant 0$. However, the integral is not zero. For example, on C, $x^2 + 4y^2 = 1$ and thus,

$$\int_C \vec{G}(x, y) \cdot d\vec{r} = \int_C \left(-y\vec{i} + x\vec{j}\right) \cdot d\vec{r} = \iint_E 2\, dx\, dy = 2\,\text{Area}(E),$$

where E is the interior of C. Thus $\vec{F}$ does not exist.

A-6. (5, 10, 2, 0, 1, 0, 0, 3, 1, 4, 80, 98)

For each integer $k \geqslant 0$, the integer n in base 3 has $k+1$ digits iff $3^k \leqslant n < 3^{k+1}-1$. Among the integers in this interval there are $\binom{k}{i}2^{k+1-i}$ for which $a(n) = i$, so

$$\sum_{n=3^k}^{3^{k+1}-1} x^{a(n)} = \sum_{i=0}^{k} \binom{k}{i} x^i 2^{k+1-i} = 2(x+2)^k.$$

Thus

$$\frac{2(x+2)^k}{3^{3k+3}} < \sum_{n=3^k}^{3^{k+1}-1} \frac{x^{a(n)}}{n^3} < \frac{2(x+2)^k}{3^{3k}},$$

and therefore

$$\frac{2}{27}\sum_{k=0}^{m}\left(\frac{x+2}{27}\right)^k < \sum_{n=1}^{3^{m+1}-1}\frac{x^{a(n)}}{n^3} < 2\sum_{k=0}^{m}\left(\frac{x+2}{27}\right)^k.$$

It follows that the series converges (for $x > 0$) iff $(x+2)/27 < 1$; that is, $0 < x < 25$.

B-1. (148, 5, 4, 1, 0, 0, 0, 0, 1, 0, 15, 30)

Solution 1. Let I denote the value of the integral. The substitution $9 - x = y + 3$ gives

$$I = \int_2^4 \frac{\sqrt{\ln(y+3)}\,dy}{\sqrt{\ln(y+3)} + \sqrt{\ln(9-y)}},$$

so

$$2I = \int_2^4 \frac{\sqrt{\ln(x+3)} + \sqrt{\ln(9-x)}}{\sqrt{\ln(x+3)} + \sqrt{\ln(9-x)}}\,dx = 2, \quad \text{and} \quad I = 1.$$

Solution 2. More generally, if

$$S = \int_a^b \frac{f(x)\,dx}{f(x) + f(a+b-x)},$$

then

$$\begin{aligned} S &= \int_a^b \left(1 - \frac{f(a+b-x)\,dx}{f(x)+f(a+b-x)}\right) dx \\ &= (b-a) - \int_b^a \frac{f(t)(-dt)}{f(a+b-t)+f(t)} = (b-a) - S, \end{aligned}$$

so $S = (b-a)/2$. In this problem $f(x) = \sqrt{\ln(9-x)}$, $a = 2$, $b = 4$, so $S = 1$.

B-2. (47, 3, 4, 0, 0, 0, 0, 0, 1, 3, 62, 84)

Solution 1. Let

$$F(r,s,t) = \frac{\binom{s}{0}}{\binom{t}{r}} + \frac{\binom{s}{1}}{\binom{t}{r+1}} + \cdots + \frac{\binom{s}{s}}{\binom{t}{r+s}}.$$

Then

$$F(r,s,t) = \frac{\binom{s-1}{0}}{\binom{t}{r}} + \frac{\binom{s-1}{0}+\binom{s-1}{1}}{\binom{t}{r+1}}$$
$$+ \cdots + \frac{\binom{s-1}{s-2}+\binom{s-1}{s-1}}{\binom{t}{r+s-1}} + \frac{\binom{s-1}{s-1}}{\binom{t}{r+s}}$$
$$= F(r,s-1,t) + F(r+1,s-1,t).$$

The proof now follows easily by induction on s.

Solution 2. We find that

$$\sum_{i=0}^{s} \frac{\binom{s}{i}}{\binom{t}{r+i}} = \frac{s!r!(t-r-s)!}{t!} \sum_{i=0}^{s} \binom{r+i}{r}\binom{t-r-i}{t-r-s}.$$

Using the binomial expansion, we see that

$$\sum_{j=0}^{\infty} \binom{t-s+1+j}{t-s+1} x^j = \frac{1}{(1-x)^{t-s+2}}$$
$$= \frac{1}{(1-x)^{r+1}} \frac{1}{(1-x)^{t-r-s+1}}$$
$$= \left(\sum_{a=0}^{\infty} \binom{r+a}{r} x^a\right)\left(\sum_{b=0}^{\infty} \binom{t-r-s-b}{t-r-s} x^b\right)$$
$$= \sum_{j=0}^{\infty} \left(\sum_{i=0}^{j} \binom{r+i}{r}\binom{t-r-s+j-i}{t-r-s}\right) x^j.$$

Equating coefficients of the x^s term, we obtain

$$\binom{t+1}{t-s+1} = \sum_{i=0}^{s} \binom{r+i}{r}\binom{t-r-i}{t-r-s}.$$

Hence

$$\sum_{i=0}^{s} \frac{\binom{s}{i}}{\binom{t}{r+i}} = \frac{s!r!(t-r-s)!}{t!} \cdot \frac{(t+1)!}{s!(t-s+1)!} = \frac{t+1}{(t-s+1)\binom{t-s}{r}}.$$

Solution 3. The problem is equivalent to proving that

$$\sum_{i=0}^{s} \frac{\binom{s}{i}\binom{t-s}{r}}{\binom{t}{r+i}} = \frac{t+1}{t+1-s}.$$

It is straightforward to show that

$$\frac{\binom{s}{i}\binom{t-s}{r}}{\binom{t}{r+i}} = \frac{t+1}{t+1-s} = \frac{\binom{r+i}{r}\binom{t-r-i}{t-s-r}}{\binom{t}{s}},$$

and, therefore, letting $l = r + i$, we have

$$\sum_{i=0}^{s} \frac{\binom{s}{i}\binom{t-s}{r}}{\binom{t}{r+i}} = \sum_{i=0}^{s} \frac{\binom{r+i}{r}\binom{t-r-i}{t-s-r}}{\binom{t}{s}} = \frac{1}{\binom{t}{s}} \sum_{l=r}^{r+s} \binom{l}{r}\binom{t-l}{t-(s+r)}.$$

Now, $\sum_{l=r}^{s+r}\binom{l}{r}\binom{t-l}{t-(s+r)}$ is the number of ways to choose a sequence of t binary digits, interrupted immediately after the lth digit by a comma, such that there are r 1's preceding the comma (among the first l digits) and $t - (s + r)$ 1's following the comma (the final $t - l$ digits).

But, interpreting the comma as another symbol in the sequence, we see that this is the number of sequences of $t + 1$ symbols containing s 0's, where the comma must be the $(r + 1)$st nonzero symbol. Thus, the sequences are completely determined by the positions of the 0's, and so the numbers of sequences is $\binom{t+1}{s}$. Thus,

$$\sum_{l=r}^{s+r} \binom{l}{r}\binom{t-l}{t-(s+r)} = \binom{t+1}{s}$$

The result follows.

B-3. (27, 31, 5, 0, 0, 0, 0, 0, 3, 11, 57, 70)
Let

$$x_r = \frac{r^2-1}{r^2+1} \quad \text{and} \quad y_r = \frac{2r}{r^2+1}$$

for r any element of F such that $r^2 \neq -1$.

It is easy to check that $(1,0)$ and (x_r, y_r) satisfy $x^2 + y^2 = 1$.

The problem is thus to show that if x, y are in F with $(x, y) \neq (1,0)$ and $x^2 + y^2 = 1$, then $x = x_r$ and $y = y_r$ for some r. We observe that for $x_r \neq 1$, $r = y_r/(1 - x_r)$, and this suggests that we set $r = y/(1 - x)$. Then we have

$$r^2 + 1 = \left(\frac{y}{1-x}\right)^2 + 1 = \frac{y^2 + (1-x)^2}{(1-x)^2}$$

$$= \frac{y^2 + x^2 - 2x + 1}{(1-x)^2} = \frac{2-2x}{(1-x)^2} = \frac{2}{1-x} \neq 0,$$

and, therefore, $r^2 - 1 = 2/(1 - x) - 2 = 2x/(1 - x)$. It follows that

$$x_r = \frac{r^2-1}{r^2+1} = x \quad \text{and} \quad y_r = \frac{2r}{r^2+1} = y.$$

B-4. (13, 14, 15, 1, 0, 0, 0, 0, 11, 23, 79, 48)

Let $y_0 = \operatorname{Arc}\cos 0.8$ and $\theta_n = y_0 + y_1 + \cdots + y_n$. Then $x_{n+1} = \cos\theta_n$ and $y_{n+1} = \sin\theta_n$. Since $\sin\theta = \sin(\pi - \theta) \leqslant \pi - \theta$ for $0 \leqslant \theta \leqslant \pi$, one easily proves by induction that $0 < \theta_n \leqslant \theta_{n+1} \leqslant \pi$ for $n = 1, 2, 3, \ldots$. Hence $L = \lim_{n\to\infty}\theta_n$ exists since $\theta_0, \theta_1, \ldots$ is a monotonic bounded sequence. It follows that $\lim_{n\to\infty} y_n = 0$.

Since $\cos t$ and $\sin t$ are continuous for all real t,

$$0 = \lim_{n\to\infty} y_{n+1} = \lim_{n\to\infty} \sin\theta_n = \sin\Big(\lim_{n\to\infty}\theta_n\Big) = \sin L.$$

As $0 < L \leqslant \pi$, this implies that $L = \pi$. Then

$$\lim_{n\to\infty} x_{n+1} = \lim_{n\to\infty}\cos\theta_n = \cos\Big(\lim_{n\to\infty}\theta_n\Big) = \cos L = \cos\pi = -1.$$

B-5. (10, 6, 3, 2, 2, 0, 0, 0, 8, 3, 16, 154)

Solution. Write $M = A + iB$ where A and B are real $2n \times n$ matrices and $N = (A, B)$, a real $2n \times 2n$ matrix. If

$$w = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = u + iv$$

where u and v are real column vectors of length n, then $\operatorname{Re}(Mw) = \operatorname{Re}[(A + iB)(u + iv)] =$

$$Au - Bv = (A \quad B)\begin{pmatrix} u \\ -v \end{pmatrix} = N\begin{pmatrix} u \\ -v \end{pmatrix}$$

and so we need to show that N is invertible. Suppose that $x = (x_1, \ldots, x_{2n})$ is a real vector of length $2n$ such that $xN = O_{2n}$. Then $xA = O_n$, $xB = O_n$ and hence $xM = x(A + iB) = O_n$. Therefore, by hypothesis, $x = O_{2n}$ and hence N has an inverse.

B-6. (8, 1, 1, 1, 0, 0, 1, 2, 1, 3, 62, 124)

Solution 1. For a in S, let $2a = \varepsilon_a s_a$ where s_a is in S and $\varepsilon_a = \pm 1$. Set $M = (p^2 - 1)/2 - N$ so that M is the number of a in S such that $\varepsilon_a = -1$. If a and b are in S and $s_a = s_b$ then $a = \pm b$ and hence $a = b$. Therefore, as a runs through S, s_a runs through S as well. Hence in F,

$$2^{(p^2-1)/2}\prod_{a\in S} a = \prod_{a\in S}(\varepsilon_a s_a) = (-1)^M \prod_{a\in S} s_a = (-1)^M \prod_{a\in S} a.$$

Hence, $(-1)^M = 2^{(p^2-1)/2}$. Using Lagrange's Theorem for finite groups or the Euler Theorem or the little Fermat Theorem, one has

$$2^{(p^2-1)/2} = (2^{p-1})^{(p+1)/2} = (1)^{(p+1)/2} = 1$$

and, therefore, M is even. But $(p^2 - 1)/2$ is also even and therefore N is even.

Solution 2. Let Z_p denote the finite field of p elements $\{0, 1, 2, \ldots, p-1\}$, and the elements of F by $\{a + bx\colon a, b \in Z_p\}$.

Let $H = \{1, 2, \ldots, (p-1)/2\}$ and $S_0 = \{a + bx\colon b \in H,\ a \in Z_p, \text{ or } b = 0 \text{ and } a \in H\}$, and set $T_0 = 2S_0$. Then S_0 satisfies the conditions of the problem and $|S_0 \cap T_0|$ is even.

Observe that any other set S with the conditions of the problem can be obtained from S_0 by a succession of exchanges of the form "take α out of S_0 and replace it with $-\alpha$."

Suppose then that S is a set which satisfies the conditions of the problem, that $T = 2S$ and that $|S \cap T|$ is even. Suppose that $\alpha \in S$ and $S' = (S - \{\alpha\}) \cup \{-\alpha\}$, $T' = 2S'$.

If $\alpha/2 \in S$ and $2\alpha \in S$ then $|S' \cap T'| = |S \cap T| - 2$ (neither α nor 2α are in $S' \cap T'$). If $\alpha/2 \in S$ and $-2\alpha \in S$, then $|S' \cap T'| = |S \cap T|$ (α is not in $S' \cap T'$ but -2α is). If $-\alpha/2 \in S$ and $2\alpha \in S$, then $|S' \cap T'| = |S \cap T|$ (2α is not in $S' \cap T'$ but $-\alpha$ is), and if $-\alpha/2 \in S$ and $-2\alpha \in S$ then $|S' \cap T'| = |S \cap T| + 2$ (both $-\alpha$ and -2α are in $S' \cap T'$).

In each case the net change in the cardinality is 0 (mod 2), and the result follows by the preceding remarks.

Letters to the Editor

Editor:

Concerning The partial order of iterated exponentials, 93 (Dec 86) with notation as defined therein:

(1) the alleged "Theorem 6", stated without proof, is in error. For instance, in $(S_5, \prec)$, the elements [23415] and [23145] are not covers, since [32145] is in between them.

(2) The conjectures at the end of the paper fail. Their truth would strongly suggest that the partial order induced on S_n coincides with the Bruhat order on S_n. They do coincide for $n \leqslant 4$; but, for instance in S_5 again, [43215] and [34512] are comparable in the former, but not in the latter.

Professor John Stembridge has written a very interesting paper which explores the relationship between the two partial orders on S_n, and which reveals (1) and (2) in the process. I am greatly indebted to him for sharing his results with me.

Barry W. Brunson
Western Kentucky University

Editor:

It was with great interest that I read Stan Wagon's article, "Fourteen Proofs of a Result about Tiling a Rectangle," in the August/September, 1987, issue of the MONTHLY. I first heard of this result while a graduate student in a complex analysis class taught by Paul Cohen. He gave the proof using the double integral of the complex exponential function, which seemed to be a triumph of analysis over combinatorial reasoning. I have given this problem to my analysis classes sporadically over the years and, to my surprise, in the spring of 1987 a proof was found by an undergraduate, Tom Amoth. His approach, given below, appears to be different than the fourteen proofs in the article.

Amoth's proof that a rectangle tiled with rectangles, each of which has at least one integer side, must itself have an integer side:
Position the large rectangle so that the origin is at the lower left-corner. Then draw all diagonal lines of the form $x + y =$ integer. Let $L(T)$ be the length of these lines that lie in rectangle T, divided by $\sqrt{2}$. Then:

1. If T has an integer side, then $L(T)$ equals the area of T. This is readily shown by cutting T into rectangles of unit width and then noting that diagonal lines run the entire length of such a rectangle only once.
2. If T has its lower-left corner on a lattice point and $L(T)$ equals the area of T, then T must have an integer side. Suppose the contrary, and assume T has dimensions $m + a$ by $n + b$ where $m, n \in \mathbb{N}$ and $0 \leqslant a,\ b < 1$. We need only consider U, the upper-right corner of T having dimensions $a \times b$. If $a + b \geqslant 1$, then $L(U) = a + b - 1 = ab - (1 - a)(1 - b) < ab$, while if $a + b < 1$, then $L(U) = 0$. In either case, $L(U) < ab$, the area of U, contradiction.

Since L and area are both additive for any interior-disjoint configuration of rectangles, these two facts imply that the large rectangle has an integer side.

Bob Burton
Department of Mathematics
Oregon State University
Corvallis, Oregon

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

Is Selfduality Involutory?

BRANKO GRÜNBAUM*
University of Washington, GN-50, Seattle, WA 98195

G. C. SHEPHARD
University of East Anglia, Norwich NR4 7TJ, England

The concept of duality occurs in many parts of mathematics. Details may vary from one field to another, but the essential feature of duality is that two objects of the same kind are related in a symmetric manner. Here we shall be interested primarily in geometric duality, though similar considerations and problems arise in other contexts.

Possibly the most familiar example of duality concerns convex polyhedra. We restrict the discussion to the three-dimensional case; there are immediate generalizations to higher dimensions. Two polyhedra P_1 and P_2 are said to be **duals** of each other if there exists a bijection δ from the family of vertices and faces of P_1, to the family of faces and vertices of P_2, which reverses inclusion. The bijection δ is called a **duality map** between P_1 and P_2. In FIGURE 1 we give an example of such a map between a cube and an octahedron. We follow the convention that δ is specified by labelling each element (vertex or face) of P_1 with the same symbol as the element of P_2 into which it is mapped by δ.

If there exists a duality map δ from a polyhedron P to itself, we say that P is **selfdual**. An example of a selfdual polyhedron is given in FIGURE 2, and a duality map δ_1 is indicated in FIGURES 2(a) and (b). In the selfdual case, the range of a duality map is equal to its domain of definition, so the duality map is a permutation on the set of vertices and faces of P. In the example, the permutation corresponding to δ_1 is given in the caption to FIGURE 2. The **rank** $r(\delta)$ of a selfduality map δ is defined as its order (period), that is, as the smallest positive integer n such that δ^n is the identity. Thus for the map δ_1 defined above we have $r(\delta_1) = 8$. However, the same polyhedron admits another selfduality map δ_2 indicated in FIGURES 2(a) and (c) which is of rank 2, that is, it is an involution. This suggests the following definition of the **rank** $r(P)$ of a selfdual polyhedron P: it is the minimum value of $r(\delta)$ over all selfduality maps δ of P.

*The last stages of this research were supported in part by National Science Foundation grant DMS-8620181.

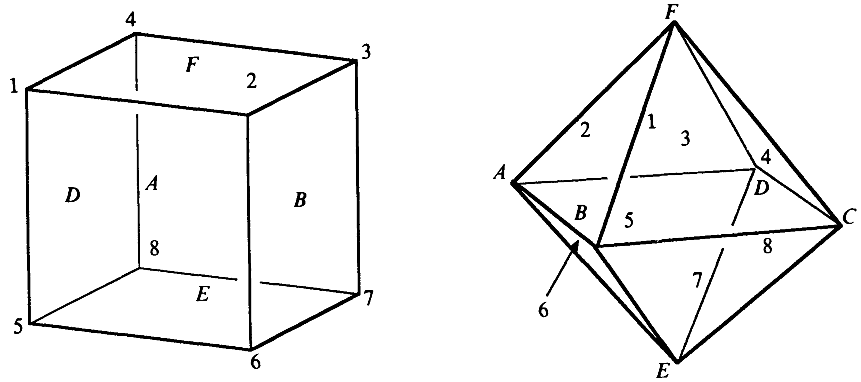


FIG. 1. Sketches of a dual pair of convex polyhedra. Vertices and faces corresponding to each other have been marked with the same label.

Problem 1. Does every selfdual convex polyhedron P have rank 2? In other words, does every selfdual P admit an involutory selfduality map?

If δ is a selfduality map, then δ^2 clearly induces a selfisomorphism of P. Hence the problem can be restated as asking whether δ can always be chosen so that this selfisomorphism is the identity. It is worth noting that if P admits a selfduality map of rank $2k$ where k is odd, then δ^k is of rank 2, so for such a polyhedron P an involutory selfduality map necessarily exists.

If a selfdual polyhedron P has rank 2 then we may use the same labels for the vertices and facets, which makes the special character of such polyhedra much more evident (see FIGURE 2(d)).

The definition of duality given above belongs to the combinatorial theory of convex polyhedra; similar considerations in the metric theory of polyhedra arise through the geometric transformation known as polarity. Assuming that the origin O is an interior point of P, the **polar** $P^* = \pi(P)$ is defined to be the convex polyhedron

$$P^* = \{x \in E^3 | \langle x, y \rangle \leqslant 1 \text{ for all } y \in P\}.$$

To each vertex v of P corresponds a face of P^*, namely that which lies in the plane with equation $\langle x, v \rangle = 1$, and to every face of P corresponds a vertex of P^*. This correspondence is easily seen to be a duality map (see, for example, Grünbaum [3, Section 3.4]). We shall say that a convex polyhedron P is **selfpolar** if P^* is the image of P under a projective map τ, and then we shall call the mapping $\sigma = \tau^{-1}\pi$ a **selfpolarity map** of P. Clearly a selfpolarity map σ induces a selfduality map δ of P, and the rank of σ is defined as the rank of the corresponding map δ. It can be shown that the rank of σ does not depend on the choice of the origin O.

Problem 2. Does every selfpolar convex polyhedron possess an involutory selfpolarity map?

An affirmative answer to this problem would imply the same answer to Problem 1 for selfpolar polyhedra—but not for selfdual polyhedra in general, since the following question is still open.

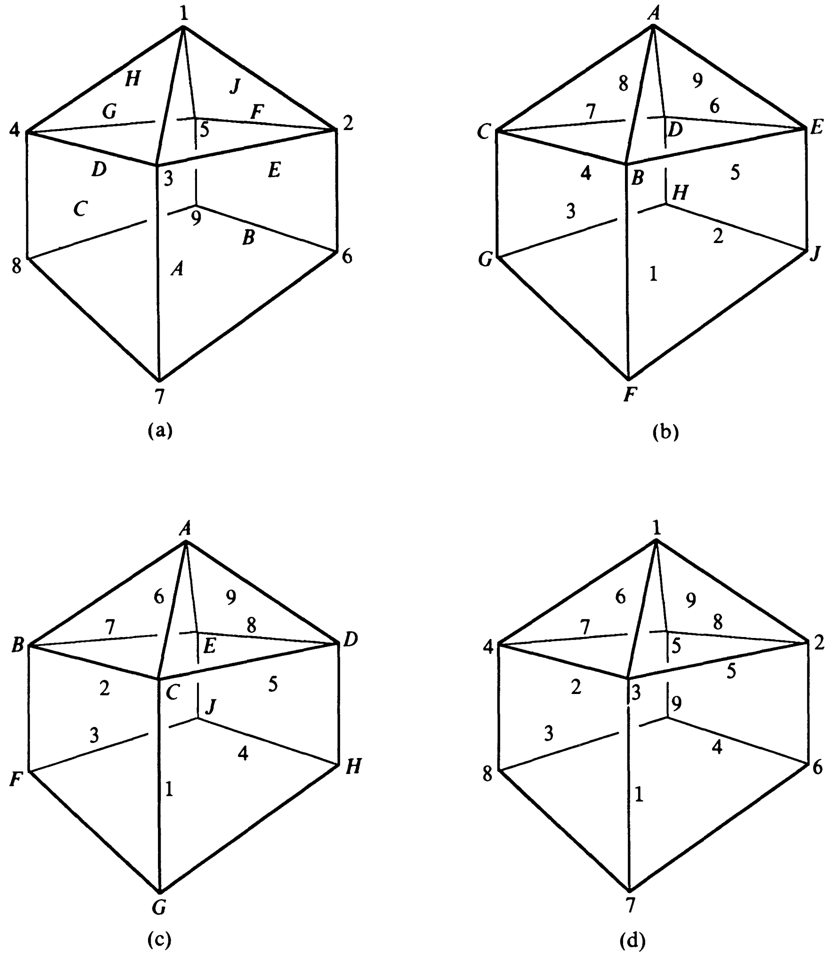


FIG. 2. (a) A convex polyhedron P. (b) A selfduality map δ_1 which shows that P is selfdual; δ_1 can be represented by the permutation (1A)(2E5D4C3B)(6J9H8G7F), hence $r(\delta_1) = 8$. (c) Another selfduality map δ_2 of P; δ_2 can be represented by the permutation (1A)(2D)(3C)(4B)(5E)(6H)(7G)(8F)(9J), hence $r(\delta_2) = 2$ and also $r(P) = 2$. (d) Another representation of the selfduality δ_2 of P, in which corresponding vertices and faces have been labelled by the same symbols.

Problem 3. Is every selfdual convex polyhedron P isomorphic to a selfpolar one?

Analogous problems arise in another example of geometric duality, namely that of configurations of points and lines in the (Euclidean or projective) plane. Here, following tradition, we say that C is an (n_k) **configuration** if C consists of a set of n distinct points and n distinct lines, such that each point of C lies on precisely k of the lines, and each line of C passes through exactly k of the points. A configuration is **selfdual** if there is an inclusion-reversing map δ of the family of points and lines of C onto itself. Examples of selfdual configurations are shown in FIGURE 3, where the points of the configurations are indicated by solid dots in order to distinguish them from those intersection points of the configuration lines which are not points of the configurations. It is well known (see, for example, Schroeter [7], Dorwart [2])

that there exist three distinct configurations (9_3) and nine configurations (10_3), each of which is selfdual. (This fact may have led to the misguided terminology of some writers who call every configuration (n_3) selfdual; our definition coincides with that which is generally accepted—see, for example, Steinitz [**8**].) The rank of a selfduality map, and the rank of a configuration, can be defined in complete analogy to the case of polyhedra. The configurations in FIGURE 3 have rank 2, as indicated by the labels, and the same is true for all the other selfdual configurations we examined, including all configurations (9_3) and (10_3).

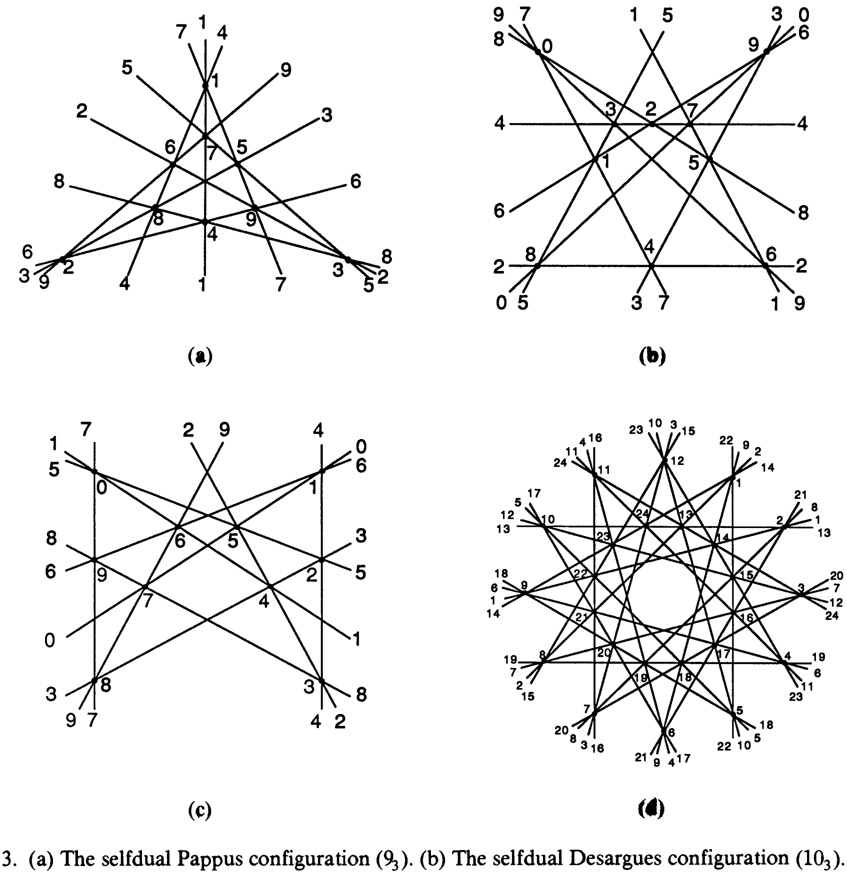


FIG. 3. (a) The selfdual Pappus configuration (9_3). (b) The selfdual Desargues configuration (10_3). (c) Another selfdual configuration (10_3). (d) A selfdual configuration (24_4).

Problem 4. Does every selfdual configuration have rank 2?

We conjecture that both Problems 1 and 4 have affirmative answers. It is clear that analogous questions can be asked in other areas of mathematics in which a duality relation can be defined, for example in the case of balanced designs.

There is a geometric counterpart to Problem 4. Let the origin be any point of the plane that does not lie on a line of the (n_k) configuration C in E^2. For each point v of C construct the line

$$\{x \in E^2 | \langle x, v\rangle = 1\},$$

and for each line $\langle x, s\rangle = 1$ of C construct the point s. The set of points and lines so obtained forms a configuration $C^* = \pi(C)$ of type (n_k) called the **polar** of C. The configuration C is called **selfpolar** if C^* is the image of C under a projective transformation τ. Then the mapping $\sigma\colon C \to C$ defined by $\sigma = \tau^{-1}\pi$ is a **selfpolarity map** of C.

Problem 5. Does every selfpolar configuration possess a selfpolarity map of rank 2?

An affirmative answer to Problem 5 would imply the same for Problem 4 if it could be established that every selfdual configuration is isomorphic to one that is selfpolar.

If the answer to Problem 1 is negative, the following more detailed question can be asked:

Problem 6. Does there exist, for every $k = 2, 3, \dots,$ a selfpolar polyhedron P_k such that $\min r(\delta) = 2k$, the minimum being taken over all selfpolarities δ of P?

Clearly, analogous problems can be posed if the answer to any of Problems 2, 4, or 5 is negative.

Since selfdual polyhedra and configurations have been under active consideration for more than a century, it is strange that problems such as those stated here not only remain unsolved, but do not even seem to have been previously asked. We can only suggest that this may be due to a confusion of thought. In all accounts we have been able to locate, the definition of selfduality for convex polyhedra is immediately followed by a completely unjustified statement which, in essence, asserts that "obviously, a polyhedron P is selfdual is another way of saying that $r(P) = 2$" (the precise wording depends, naturally, on each author's terminology). As examples of this substitution of meaning we may mention Brückner [**1**, p. 76, lines -2, -1], Hermes [**4**, p. 132, line 17], Jucovič [**5**, p. 5, lines 3–6]; in the presentation due to A. Cayley (see Kirkman [**6**, p. 185]) the precise meaning of "selfdual" ("autopolar") is not made clear, but the notation indicates that the same idea was intended.

REFERENCES

1. M. Brückner, Vielecke und Vielflache, Teubner, Leipzig, 1900.
2. Dorwart, Configurations, WFF'N PROOF, New Haven, CT, 1967.
3. B. Grünbaum, Convex Polytopes, Interscience, London, 1967.
4. O. Hermes, Die Formen der Vielflache, *J. reine angew. Math.*, 122 (1900) 124–154.
5. E. Jucovič, Selfconjugate *K*-polyhedra [in Russian, with German summary]. *Mat.-Fyz. Časopis Sloven. Akad. Vied*, 12 (1962) 1–22.
6. T. P. Kirkman, On autopolar polyedra, *Philos. Trans. Roy. Soc. London*, 147 (1857) 183–215.
7. H. Schroeter, Ueber die Bildungsweise und geometrische Konstruction der Konfigurationen (10_3), *Nachr. königl. Ges. Wiss. und der Georg-Augusts-Universität zu Göttingen*, No. 8 (1889) 193–236.
8. E. Steinitz, Über Konfigurationen, *Archiv Math. und Physik*, (3) 16 (1910) 289–313.

NOTES

EDITED BY DENNIS DETURCK, DAVID J. HALLENBECK, AND ANITA E. SOLOW

A Simple Verification of Ilieff's Conjecture for Polynomials With Three Zeros

G. L. COHEN AND G. H. SMITH

School of Mathematical Sciences, University of Technology, Sydney, New South Wales 2007, Australia

Ilieff's Conjecture has been discussed by Marden [1], [2], who in fact attributes the conjecture to the Bulgarian mathematician B. Sendov in 1962. It states: *If $P(z)$ is any polynomial with all its zeros in $|z| \leqslant 1$, and z_0 is any one such zero, then $P'(z)$ has a zero in $|z - z_0| \leqslant 1$.*

Proofs of this have been given for polynomials of degree at most five and in a number of other special cases for polynomials of any degree. References to these results may be found in Marden [1] and Schmeisser [4]. Five or six different methods of verification of Ilieff's Conjecture have been given for cubics, and two (by Saff and Twomey [3] and Schmeisser [4]) also include the result for polynomials $P(z) = (z - z_0)^{n_0}(z - z_1)^{n_1}(z - z_2)^{n_2}$, considered here. Our justification for giving another proof is that our approach is fully self-contained and extremely simple.

The result is quickly achieved for polynomials $P(z)$ as just given (where, by hypothesis, z_0, z_1, z_2 lie in $|z| \leqslant 1$), when $n = n_0 + n_1 + n_2 \geqslant 4$. We set

$$P'(z) = n(z - z_0)^{n_0 - 1}(z - z_1)^{n_1 - 1}(z - z_2)^{n_2 - 1}(z - w_1)(z - w_2), \tag{1}$$

and wish to show that $P'(z)$ has a zero in $|z - z_0| \leqslant 1$. This is immediate if $n_0 > 1$, so suppose $n_0 = 1$. Writing $P(z) = (z - z_0)Q(z)$, we see that

$$P'(z_0) = Q(z_0) = (z_0 - z_1)^{n_1}(z_0 - z_2)^{n_2}. \tag{2}$$

From (1), with $n_0 = 1$, and (2),

$$n|z_0 - w_1||z_0 - w_2| = |z_0 - z_1||z_0 - z_2| \leqslant 4. \tag{3}$$

Hence, if $n \geqslant 4$, then $|z_0 - w_1||z_0 - w_2| \leqslant 1$, so that $|z_0 - w_j| \leqslant 1$ for $j = 1$ or 2. Thus Ilieff's Conjecture is verified when $P(z)$ has a multiple zero.

Remark. It follows in the same way that *Ilieff's Conjecture is true for any polynomial with p zeros if its degree n satisfies $n \geqslant 2^{p-1}$.*

In the only case remaining for the polynomial $P(z)$ above, we have $n_0 = n_1 = n_2 = 1$. For this, we use the general

LEMMA. *Let $P(z)$ be any polynomial of degree $n \geqslant 2$. Let z_0 be a simple zero of $P(z)$ and let $w_1, w_2, \ldots, w_{n-1}$ be the zeros of $P'(z)$. If $|P''(z_0)| \geqslant (n-1)|P'(z_0)|$, then $|z_0 - w_j| \leqslant 1$ for at least one $j = 1, 2, \ldots, n-1$.*

Proof. We may assume $P(z)$ is monic. Then $P'(z) = n\Pi_{j=1}^{n-1}(z - w_j)$. If $P'(z) \neq 0$, we have, on taking logarithms and differentiating,

$$\frac{P''(z)}{P'(z)} = \sum_{j=1}^{n-1} \frac{1}{z - w_j}.$$

Since z_0 is a simple zero of $P(z)$, $z_0 \neq w_j$ for any j. Suppose $|P''(z_0)| \geqslant (n-1)|P'(z_0)|$. Then if $|z_0 - w_j| > 1$ for all j, we have

$$n - 1 \leqslant \left|\frac{P''(z_0)}{P'(z_0)}\right| \leqslant \sum_{j=1}^{n-1} \frac{1}{|z_0 - w_j|} < n - 1,$$

a contradiction. Hence $|z_0 - w_j| \leqslant 1$ for some j.

Notice also that, for the same polynomial with zeros $z_0, z_1, \dots, z_{n-1}$, we may write $P(z) = (z - z_0)Q(z)$ and obtain

$$\frac{P''(z_0)}{P'(z_0)} = 2\frac{Q'(z_0)}{Q(z_0)} = 2\sum_{j=1}^{n-1} \frac{1}{z_0 - z_j}. \tag{4}$$

Returning now to the cubic $P(z) = (z - z_0)(z - z_1)(z - z_2)$, we set $x = |z_0 - z_1|$, $y = |z_0 - z_2|$, $a = |z_0 - \frac{1}{2}(z_1 + z_2)|$, $b = \frac{1}{2}|z_1 - z_2|$, so $0 < x \leqslant 2$, $0 < y \leqslant 2$, $0 \leqslant a \leqslant 2$, $0 \leqslant b \leqslant 1$. See FIGURE 1 in which the vertices of the triangle ABC are at z_0, z_1, z_2, respectively.

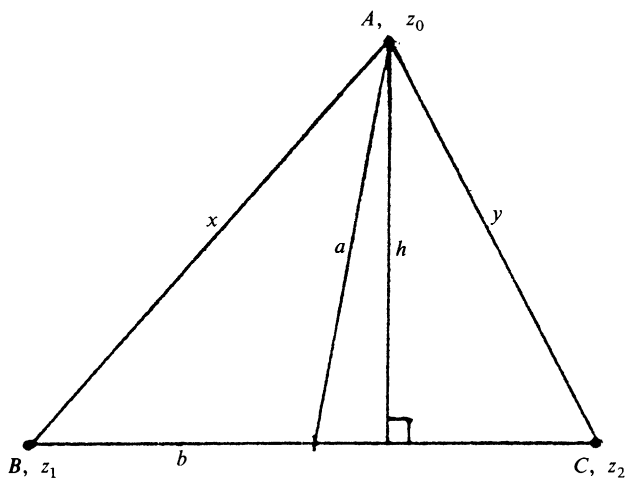


FIG. 1.

From (4), we have

$$\left|\frac{P''(z_0)}{P'(z_0)}\right| = 2\left|\frac{1}{z_0 - z_1} + \frac{1}{z_0 - z_2}\right| = \frac{4a}{xy},$$

so, by the Lemma, Ilieff's Conjecture is true whenever $xy \leqslant 2a$, and in particular (since $x \leqslant 2$ and $y \leqslant 2$) whenever $x \leqslant a$ or $y \leqslant a$. We assume therefore that $x > a$ and $y > a$.

From (3), we have $3|z_0 - w_1||z_0 - w_2| = xy$, so Ilieff's Conjecture is true if $xy \leqslant 3$. Assuming then that $xy > 3$, we have $x^2 + y^2 = (x - y)^2 + 2xy > 6$. Since

$x > a$, $y > a$ and $\cos \hat{A} = (x^2 + y^2 - 4b^2)/2xy > (6 - 4)/2xy = 1/xy > 0$, triangle ABC in FIGURE 1 is acute-angled. Then the triangle's circumradius R satisfies $R \leqslant 1$ (proved below) and, if h is the length of the perpendicular from A to BC, a classical theorem in Euclidean geometry (also proved below) gives us

$$xy = 2Rh \leqslant 2a.$$

As above, the Lemma confirms Ilieff's Conjecture in this case, and our proof is complete.

Two matters remain to be clarified.

In FIGURE 2, O is the circumcentre of the triangle ABC, E is on the circumcircle and D is the foot of the perpendicular from A to BC. The right-angled triangles ABD, AEC are similar, so $|AB|/|AD| = |AE|/|AC|$, from which, in the notation above, $xy = 2Rh$.

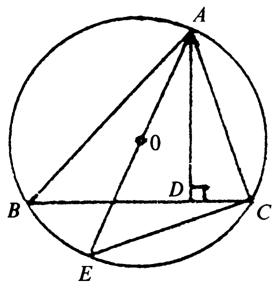


FIG. 2

We are not sure whether it is "obvious" or not, that the circumradius R of an acute-angled triangle lying within a unit circle satisfies $R \leqslant 1$. The following is a proof. Let ABC be the triangle. If A, B, C are on the unit circle then $R = 1$. If B, C are on the unit circle and A is not then, since the triangle is acute-angled, A is in the major segment, and using FIGURE 3 in which F is the midpoint of BC and A' is on the unit circle,

$$2R = \frac{|BC|}{\sin \hat{A}} < \frac{|BC|}{\sin \hat{A}'} = 2.$$

By rotation and/or translation, perhaps renaming the vertices, the other possibilities may be reduced to this one.

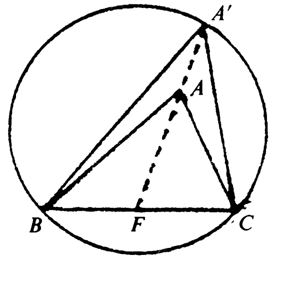


FIG. 3

REFERENCES

1. Morris Marden, Conjectures on the critical points of a polynomial, this MONTHLY, 90 (1983) 267–276.
2. ______, The search for a Rolle's Theorem in the complex domain, this MONTHLY, 92 (1985) 643–650.
3. E. B. Saff and J. B. Twomey, A note on the location of critical points of polynomials, *Proc. Amer. Math. Soc.*, 27 (1971) 303–308.
4. Gerhard Schmeisser, On Ilieff's conjecture, *Math. Z.*, 156 (1977) 165–173.

UC Spaces Revisited

GERALD BEER
Department of Mathematics, California State University, Los Angeles, CA 90032

In any exposure to metric space topology, spaces with the following properties are considered: (a) each continuous function on the space with values in an arbitrary metric space is uniformly continuous; (b) each open cover of the space has a *Lebesgue number*, i.e., there exists a number $\delta > 0$ such that each subset of diameter less than δ lies entirely in some member of the cover. Both of these properties hold for compact metric spaces, but as is well known, neither characterizes compactness, and the class of spaces in which (a) holds is precisely the class in which (b) holds. This larger class of spaces has been called the *UC spaces* [**11**], the *Lebesgue spaces* [**7**], or the *Atsuji spaces* ([**2**], [**3**]) (in recognition of Atsuji's fundamental paper [**1**]). This class seems to be periodically rediscovered in this MONTHLY ([**4**] and [**6**]). Perhaps the most visual internal characterization of a UC space X is the following [**6**]: the set X' of limit points of X is compact, and for each $\varepsilon > 0$, $\{x \in X\colon d(x, X') \geqslant \varepsilon\}$ is *uniformly discrete*, i.e., there exists $\delta > 0$ such that if $d(x, X') \geqslant \varepsilon$ and $d(y, X') \geqslant \varepsilon$ and $x \neq y$, then $d(x, y) > \delta$. From this description, one can easily construct exotic UC spaces. For example, if $\{e_n\colon n \in Z^+\}$ is the standard orthonormal base in l_2, then

$$\{(1/j)e_n\colon n \in Z^+,\ j \in Z^+\} \cup \{0\}$$

is a nonlocally compact UC space.

Among the many characterizations of UC spaces, perhaps the most intriguing describe the way they sit between the compact spaces and the complete ones. A space X is compact if and only if each sequence with distinct terms has a cluster point, and it is complete if and only if each sequence with distinct terms *in which pairs of terms are arbitrarily close eventually* has a cluster point. Similarly, a space X is UC if and only if each sequence with distinct terms *in which pairs of terms are arbitrarily close frequently* has a cluster point ([**3**] and [**10**]). With respect to decreasing sequences of nonempty closed sets, a space X is compact if and only if each such sequence has nonempty intersection, and a space X is complete if and only if each such sequence *whose diameters go to zero* has nonempty intersection. An analogous characterization holds for UC spaces [**3**], provided we replace the diameter functional by another geometric functional λ, defined as follows:

$$\lambda(\varnothing) = 0$$

$$\lambda(A) = \sup\{d(a, X - \{a\})\colon a \in A\} \qquad (A \neq \varnothing).$$

Intuitively, λ measures the maximal degree to which points of a subset A of the space X are isolated.

Which metrizable spaces X admit a UC metric? The answer is quite simple: X' must be compact. In essence, this result is due to Nagata [**8**], who showed that the fine uniformity for a metrizable space is a metric uniformity precisely when X' is compact (see also [**9**]). The result also follows from a nontrivial fact about refinements of sequences of open covers of a metrizable space X: if $\langle \Omega_n \rangle$ is a sequence of open covers of X, then there exists a compatible metric ρ such that for each positive integer n, $\{\{w \in X: \rho(x, w) < 1/n\}: x \in X\}$ refines Ω_n [**5**; p. 196]. To see this, let d be any compatible metric for X. For convenience, we denote the *parallel body* $\{x \in X: d(x, X') < \varepsilon\}$ by $S_d[X'; \varepsilon]$. For each $n \in Z^+$, let Ω_n be the following open cover of X:

$$\Omega_n = \{S_d[X'; 1/n]\} \cup \{\{x\}: x \in X - X'\}.$$

Now let ρ be the metric induced by the sequence $\langle \Omega_n \rangle$ as described above. To show that ρ is a UC metric, we need only show that for each $\varepsilon > 0$, $(S_\rho[X'; \varepsilon])^c$ is uniformly discrete. If not, then for some positive ε, we can construct a sequence

$$x_1, w_1, x_2, w_2, x_3, w_3, \ldots$$

in $(S_\rho[X'; \varepsilon])^c$ such that for each n, $0 < \rho(x_n, w_n) < 1/n$. By construction, for each $n > 1/\varepsilon$, x_n must lie in $S_d[X'; 1/n]$, for otherwise, the ρ-ball of radius $1/n$ about x_n would contain no other point (in particular, it would not contain w_n). As a result, $d(x_n, X') \to 0$; so, by the compactness of X', $\langle x_n \rangle$ has a cluster point p. Since all terms of the sequence are isolated points, no term can occur infinitely often, so, the point p must be a limit point of X. This produces a contradiction, for our sequence lies entirely in the closed set $(S_\rho[X'; \varepsilon])^c$.

Neither approach above to the question of *UC* metrizability seems appropriate for the nontopologist. We produce the first utterly elementary proof.

THEOREM. *Let X be a metrizable space. If X' is compact, then X is UC metrizable.*

Proof. If $X' = \varnothing$, then the zero-one metric is compatible with the (discrete) topology of the space. Thus, in this case, X is UC metrizable. Otherwise let d be a compatible metric for X. Define ρ: $X \times X \to [0, \infty)$ by

$$\rho(x, y) = \begin{cases} 0 & \text{if } x = y \\ d(x, y) + \max\{d(x, X'), d(y, X')\} & \text{if } x \neq y. \end{cases}$$

Now if α, β, and γ are nonnegative, we have

$$\max\{\alpha, \gamma\} \leqslant \alpha + \gamma \leqslant \max\{\alpha, \beta\} + \max\{\beta, \gamma\}$$

Since ρ is the sum of a metric and a pseudometric, ρ is a metric. Clearly, ρ is at least as strong as d. Conversely, we show each ρ-ball contains a concentric d-ball. To this end, fix $x \in X$ and $\varepsilon > 0$. If $x \in X - X'$, there exists $\delta > 0$ with $S_d[x; \delta] = \{x\}$. Thus, $S_d[x; \delta] \subset S_\rho[x; \varepsilon]$. Otherwise, $x \in X'$. If $d(x, y) < \varepsilon/2$, then

$$\begin{aligned} \rho(x, y) &= d(x, y) + \max\{d(x, X'), d(y, X')\} \\ &< \varepsilon/2 + \max\{0, d(y, x)\} < \varepsilon/2 + \varepsilon/2 = \varepsilon. \end{aligned}$$

Thus, $S_d[x; \varepsilon/2] \subset S_\rho[x; \varepsilon]$.

To show that ρ is a UC metric, we again need only show that for each $\varepsilon > 0$, $(S_\rho[X'; \varepsilon])^c$ is uniformly discrete. If not, then just as before, for some positive ε, we can construct a sequence

$$x_1, w_1, x_2, w_2, x_3, w_3, \ldots$$

in $(S_\rho[X'; \varepsilon])^c$ such that for each n, $0 < \rho(x_n, w_n) < 1/n$. By the definition of the metric ρ, for each n, $d(x_n, X') < 1/n$. By the compactness of X', $\langle x_n \rangle$ has a cluster point in X', leading us to the desired contradiction. ■

We remark in closing that if $\langle X, d_X \rangle$ and $\langle Y, d_Y \rangle$ are arbitrary metric spaces, and f: $X \to Y$ is continuous, then there always exists an equivalent metric for X which makes that single function uniformly continuous, namely,

$$\rho(x, w) = d_X(x, w) + d_Y(f(x), f(w)).$$

REFERENCES

1. M. Atsuji, Uniform continuity of continuous functions of metric spaces, *Pacific J. Math.*, 8 (1958) 11–16.
2. G. Beer, Metric spaces on which continuous functions are uniformly continuous and Hausdorff distance, *Proc. Amer. Math. Soc.*, 95 (1985) 653–658.
3. ______, More about metric spaces on which continuous functions are uniformly continuous. *Bull. Australian Math. Soc.*, 33 (1986) 397–406.
4. M. A. Chaves, Spaces where all continuity is uniform, *Amer. Math. Monthly*, 92 (1985) 487–489.
5. J. Dugundji, Topology, Allyn and Bacon, Boston, 1966.
6. H. Hueber, On uniform continuity and compactness in metric spaces, *Amer. Math. Monthly*, 88 (1981) 204–205.
7 S. Nadler and T. West, A note on Lebesgue spaces, *Topology Proc.*, 6 (1981) 363–369.
8. J. Nagata, On the uniform topology of bicompactifications, *J. Inst. Polytech. Osaka*, 1 (1950) 28–38.
9. J. Rainwater, Spaces whose finest uniformity is metric, *Pacific J. Math.*, 9 (1959) 567–570.
10. G. Toader, On a problem of Nagata, *Mathematica (Cluj)*, 20 (43) (1978) 78–79.
11. W. Waterhouse, On UC Spaces, *Amer. Math. Monthly*, 72 (1965) 634–635.

Line Integrals of Higher-Order Derivatives

G. N. HILE AND R. Z. YEH
Mathematics Department, University of Hawaii, Honolulu, Hawaii 96822

1. Introduction. In advanced calculus one learns that a smooth function can be recovered from its gradient via a line integral; indeed, one has the formula

$$f(x) = f(x_0) + \int_{x_0}^{x} \nabla f(y) \cdot dy, \tag{1}$$

where the line integral can be taken over any sufficiently regular curve running from x_0 to x. But what if only higher order derivatives of f are known, say all partial derivatives of order m for some integer m greater than 1? It seems to be not generally recognized that an analogue of (1) holds for higher order derivatives; one can recover f from its derivatives of order m via a line integral. Uniqueness is guaranteed up to an arbitrary polynomial of degree $m - 1$. We give the formula and

explain the notation later; it is

$$f(x) = \sum_{|\alpha| \leqslant m-1} \frac{D^\alpha f(0)}{\alpha!} x^\alpha - \int_0^x \sum_{|\alpha|=m} \frac{D^\alpha f(y)}{\alpha!} d\left[(x-y)^\alpha\right], \tag{2}$$

where for convenience we take $x_0 = 0$ as the starting point for the line integral.

A related problem in advanced calculus is that of determining whether an n-tuple $(u_1, \ldots, u_n)$ of functions in a domain in $\mathbb{R}^n$ is the gradient of some function f. If the domain is simply connected and the functions $\{u_i\}$ are of class C^1, the so-called compatibility conditions,

$$\partial u_i/\partial x_j = \partial u_j/\partial x_i, \qquad 1 \leqslant i < j \leqslant n, \tag{3}$$

are necessary and sufficient to guarantee the existence of such an f. The same question can be raised for higher order derivatives, with an analogous answer. If certain compatibility conditions are met, then a given collection of functions $\{u_\alpha\}$ will correspond to the derivatives of order m of some function f.

2. Recovery of a Function from Derivatives. It is important to use economical notation; otherwise, one is quickly lost in a sea of subscripts. We let $\alpha = (\alpha_1, \ldots, \alpha_n)$ denote a *multiindex* of dimension n; thus α is an n-tuple whose components are nonnegative integers. The differential operator D^α, as customarily defined, is

$$D^\alpha := \partial_1^{\alpha_1} \partial_2^{\alpha_2} \cdots \partial_n^{\alpha_n},$$

where ∂_i denotes differentiation with respect to the ith variable. The *order* of D^α is the quantity

$$|\alpha| := \alpha_1 + \alpha_2 + \cdots + \alpha_n.$$

Thus if f is a real-valued function of $x = (x_1, x_2, \ldots, x_n)$, $D^\alpha f$ is the derivative

$$D^\alpha f := \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \cdots \partial x_n^{\alpha_n}}.$$

We always assume that our functions are sufficiently smooth so that mixed partials do not depend on the order of differentiation; for example, $\partial^2 f/\partial x_i \partial x_j = \partial^2 f/\partial x_j \partial x_i$. How many distinct derivatives does f have of order m? The question is equivalent to asking in how many ways m indistinguishable apples can be distributed among n people; hence ([2], [3]) the formula is $\binom{m+n-1}{m}$.

Monomials x^α are defined as

$$x^\alpha := x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n},$$

and the generalized factorial $\alpha!$ is

$$\alpha! := \alpha_1! \alpha_2! \cdots \alpha_n!.$$

A *domain* $\mathscr{D}$ in $\mathbb{R}^n$ is a connected open set. A *smooth curve* $\mathscr{C}$ in $\mathscr{D}$ is a parametrized set $\mathscr{C} = \{y(t)\colon a \leqslant t \leqslant b\}$, such that y is a C^1 function from $[a, b]$ into $\mathscr{D}$, with velocity $y'(t)$ never zero. A *simple* smooth curve never meets itself except possibly at the two ends. For real-valued functions f in $C(\mathscr{D})$ and g in $C^1(\mathscr{D})$, we have the line integral

$$\int_{\mathscr{C}} f(y)\, dg(y) := \int_{\mathscr{C}} f(y) \nabla g(y) \cdot dy := \int_{\mathscr{C}} \sum_{i=1}^n f(y) \frac{\partial g(y)}{\partial y_i} dy_i.$$

For f and g both in $C^1(\mathscr{D})$, it is not difficult to verify the integration by parts formula,

$$\int_{\mathscr{C}} f\,dg + \int_{\mathscr{C}} g\,df = \int_{\mathscr{C}} d(fg) = f(y)g(y)\Big|_{y=y(a)}^{y=y(b)}.$$

We now verify the recovery formula (2) under the assumptions that f is a real valued function in $C^m(\mathscr{D})$, with the origin 0 lying in $\mathscr{D}$, and with the line integral taken over any simple smooth curve in $\mathscr{D}$ running from 0 to x. For the case $m = 1$ the formula reduces to (1) with $x_0 = 0$, and so the result is already known. We assume the result is valid for a given positive integer m, and verify it for the next integer $m + 1$.

Using integration by parts we can verify the identity

$$-\int_0^x \frac{D^\alpha f(y)}{\alpha!} d\left[(x-y)^\alpha\right] = \frac{D^\alpha f(0)}{\alpha!} x^\alpha + \int_0^x \frac{(x-y)^\alpha}{\alpha!} d\left[D^\alpha f(y)\right];$$

thus (2) may be written also as

$$f(x) = \sum_{|\alpha| \leqslant m} \frac{D^\alpha f(0)}{\alpha!} x^\alpha + \int_0^x \sum_{|\alpha|=m} \frac{(x-y)^\alpha}{\alpha!} d\left[D^\alpha f(y)\right]. \tag{4}$$

Let $e^i = (0, \dots, 1, 0, \dots, 0)$ denote the ith unit multi-index of dimension n. We have the chain of equalities

$$\begin{aligned}\sum_{|\alpha|=m} \frac{(x-y)^\alpha}{\alpha!} d\left[D^\alpha f(y)\right] &= \sum_{|\alpha|=m} \frac{(x-y)^\alpha}{\alpha!} \sum_{i=1}^n \partial_i D^\alpha f(y)\, dy_i \\ &= \sum_{i=1}^n \sum_{|\alpha|=m} \left(-\frac{\partial}{\partial y_i} \frac{(x-y)^{\alpha+e^i}}{(\alpha+e^i)!} \right) D^{\alpha+e^i} f(y)\, dy_i \\ &= -\sum_{i=1}^n \sum_{|\beta|=m+1,\ \beta_i \geqslant 1} D^\beta f(y) \frac{\partial}{\partial y_i} \left(\frac{(x-y)^\beta}{\beta!} \right) dy_i.\end{aligned}$$

But since $\dfrac{\partial}{\partial y_i}(x-y)^\beta = 0$ if $\beta_i = 0$, the condition $\beta_i \geqslant 1$ can be discarded, and the chain continues as

$$= -\sum_{|\beta|=m+1} \frac{D^\beta f(y)}{\beta!} \sum_{i=1}^n \frac{\partial}{\partial y_i}\left[(x-y)^\beta\right] dy_i = -\sum_{|\beta|=m+1} \frac{D^\beta f(y)}{\beta!} d\left[(x-y)^\beta\right].$$

This last expression can now be substituted into (4) to yield formula (2), with m in (2) replaced by $m + 1$.

3. Compatibility Conditions. Let $\{u_\alpha\}_{|\alpha|=m}$ be a collection of real-valued functions of class C^1 in a simply connected domain $\mathscr{D}$ in $\mathbb{R}^n$. What conditions guarantee the existence of a function f of class C^{m+1} in $\mathscr{D}$ such that $u_\alpha = D^\alpha f$ for each u_α? If such an f exists, then clearly the equations

$$\partial_i u_{\beta+e^j} = \partial_j u_{\beta+e^i}, \quad \text{for } |\beta| = m-1, \qquad 1 \leqslant i < j \leqslant n, \tag{5}$$

must hold, as each side of the equality reduces to $D^{\beta+e^i+e^j} f$. For example, take

$m = 3$, $n = 2$; then we have the compatibility conditions

$$\partial_1 u_{(2,1)} = \partial_2 u_{(3,0)}, \qquad \partial_1 u_{(1,2)} = \partial_2 u_{(2,1)}, \qquad \partial_1 u_{(0,3)} = \partial_2 u_{(1,2)}.$$

In general (5) is a system of equations in which there are $\binom{m+n-2}{m-1}$ choices for β with $|\beta| = m - 1$, and then $\binom{n}{2}$ choices of i and j with $1 \leqslant i < j \leqslant n$; hence there is a total of

$$\binom{m+n-2}{m-1} \cdot \binom{n}{2}$$

compatibility equations listed in (5). We show that these equations are sufficient; if they all hold then the existence of f is guaranteed, as long as the domain $\mathscr{D}$ is simply connected. We show further that f can be obtained explicitly as a line integral involving the functions $\{u_\alpha\}$, and that f is uniquely determined up to a polynomial of degree $m - 1$.

For $m = 1$, (5) reduces to the usual conditions (3), and again the result is already known ([**1**], [**4**]). We assume the result is true for a given positive integer m, and proceed to the case $m + 1$. So assume we have real-valued functions $\{u_\alpha\}_{|\alpha| = m+1}$ all in $C^1(\mathscr{D})$, with

$$\partial_i u_{\beta + e^j} = \partial_j u_{\beta + e^i}, \quad \text{for } |\beta| = m, \qquad 1 \leqslant i < j \leqslant n.$$

These formulas allow us to apply the case $m = 1$ to each collection $u_{\beta+e^1}$, $u_{\beta+e^2}, \ldots, u_{\beta+e^n}$, to conclude that for each β with $|\beta| = m$ there exists a function u_β in $C^2(\mathscr{D})$ such that

$$\partial_i u_\beta = u_{\beta + e^i}, \quad \text{for } i = 1, \ldots, n.$$

Moreover, if $|\gamma| = m - 1$, then $|\gamma + e^i| = m$, and

$$\partial_i u_{\gamma + e^j} = u_{\gamma + e^j + e^i} = u_{\gamma + e^i + e^j} = \partial_j u_{\gamma + e^i}.$$

Thus the family $\{u_\beta\}_{|\beta| = m}$ satisfies the induction hypotheses, implying the existence of a function f in $C^{m+1}(\mathscr{D})$ such that

$$D^\beta f = u_\beta, \quad \text{for } |\beta| = m.$$

But actually f lies in $C^{m+2}(\mathscr{D})$, since each u_β is in $C^2(\mathscr{D})$. Furthermore, if $|\alpha| = m + 1$, then $\alpha = \beta + e^i$ for some β with $|\beta| = m$ and some e^i; thus

$$D^\alpha f = D^{\beta + e^i} f = \partial_i \left(D^\beta f\right) = \partial_i u_\beta = u_{\beta + e^i} = u_\alpha.$$

The induction proof is complete.

Now that we know some f exists, we see in view of (2) that any such f must be of the form

$$f(x) = \sum_{|\alpha| \leqslant m-1} \frac{D^\alpha f(0)}{\alpha!} x^\alpha - \int_0^x \sum_{|\alpha| = m} \frac{u_\alpha(y)}{\alpha!} d\left[(x - y)^\alpha\right],$$

where the first summation on the right provides a polynomial of degree $m - 1$. If we replace this polynomial by another polynomial of the same degree, we obtain a function $g(x)$ differing from $f(x)$ by a polynomial of degree $m - 1$. Then for every α with $|\alpha| = m$, we have $D^\alpha g(x) = D^\alpha f(x) = u_\alpha(x)$. This argument shows that the totality of functions having $\{u_\alpha\}$ as mth order derivatives are precisely those represented by the above equation for f.

4. An Application and Further Remarks

(a) The discovery of formula (2) arose in our study of the reduction of a higher order partial differential equation to a first order system. As a simple example, consider the biharmonic equation in the plane,

$$\Delta^2\phi = \phi_{xxxx} + 2\phi_{xxyy} + \phi_{yyyy} = 0. \tag{6}$$

We reduce this equation to a first order system by naming all the third order derivatives of ϕ;

$$u = \phi_{xxx}, \qquad v = \phi_{xxy}, \qquad p = \phi_{xyy}, \qquad q = \phi_{yyy}. \tag{7}$$

Then we obtain from (6) and differentiation of (7) the first order system

$$\begin{aligned} u_x + 2v_y + q_y &= 0 \\ v_x - u_y &= 0 \\ p_x - v_y &= 0 \\ q_x - p_y &= 0. \end{aligned} \tag{8}$$

Now assuming that we have a solution (u, v, p, q) of the system, the problem arises of recovering a solution ϕ of the biharmonic equation. The last three equations of (8) are compatibility conditions; therefore we may use formula (2) to obtain a function ϕ such that the prescriptions (7) hold. The first equation of (8) then reveals that ϕ is a solution of the biharmonic equation.

A similar analysis applies to more general higher order partial differential equations in any number of independent variables.

(b) Formula (2) can be viewed as a generalized Taylor formula with remainder. It can be used to determine sufficient conditions for the validity of the infinite expansion

$$f(x) = \sum_{\alpha} \frac{D^\alpha f(0)}{\alpha!} x^\alpha,$$

where the sum is taken over all multi-indices α of dimension n.

(c) If the origin 0 is replaced by an arbitrary point x_0 in $\mathscr{D}$, then (2) is modified by replacing 0 with x_0 and x^α with $(x - x_0)^\alpha$.

(d) Versions of (2) appear in the literature ([5], [**6**]), but with the line integral taken over a straight line running from 0 to x. They are obtained by parametrizing the line and applying the Taylor formula with remainder for functions of one variable, with f being viewed as a function of one variable along the line. Our formula is more versatile because we do not require that 0 and x can be connected by a straight line, a requirement which can be met of course only in restricted domains.

REFERENCES

1. R. C. Buck, Advanced Calculus, McGraw-Hill, New York, 1978.
2. K. L. Chung, Elementary Probability Theory with Stochastic Processes, Springer-Verlag, New York, 1974.

3. W. Feller, An Introduction to Probability Theory and Its Applications, Vol. 1, Wiley, New York, 1957.
4. W. H. Fleming, Functions of Several Variables, Addison-Wesley, Reading, 1965.
5. J. E. Marsden and A. J. Tromba, Vector Calculus, Freeman, San Francisco, 1976.
6. R. T. Seeley, Calculus of Several Variables, Scott, Foresman and Co., Glenview, Il., 1970.

A Note on Closure Continuity

JAMES CHEW
Mathematics Department, N.C. A & T State University, Greensboro, NC 27411

In their classroom note [1], Andrew and Whittlesy gave the following definitions: A function $f\colon X \to Y$ is closure continuous at $x \in X$ provided that if V is an open set in Y containing $f(x)$, there exists an open set U in X containing x such that $f(\bar{U}) \subset \bar{V}$. (Here context makes clear in which topological spaces the closure operations are taken; i.e. the bar on U refers to closure in X, while the bar on V refers to closure in Y.) The function f is said to be closure continuous on X if it is closure continuous at each point of X.

Theorem 1 of [1] states that a function that is continuous on X is closure continuous on X. An example is given showing that the converse is false. Thus global closure continuity is strictly weaker than global continuity.

The purpose of this note is to exhibit a relationship between closure continuity and a well-known generalization of continuity. Recall that a function $f\colon X \to Y$ is called connected if the image of each connected set in X is a connected set in Y.

THEOREM 1. *If $f\colon X \to Y$ is closure continuous on X, then f is connected.*

Proof. In what follows, we may assume that the given connected set in X is all of X and that $Y = f(X)$. (If not, we simply consider the relevant relative topologies.) Suppose that Y is not connected. Then $Y = V \cup W$, where V and W are nonempty, disjoint open sets. Let $A = f^{-1}(V)$. For each $x \in A$, there exists an open set U containing x such that $f(\bar{U}) \subset \bar{V}$ by closure continuity of f on X. We claim that $U \subset A$. If not, $f(\bar{U})$ would intersect W, say at p. Then $p = f(x')$ for some $x' \in U$ and so $f(x') \in \bar{V} = V$ or $p \in V \cap W$, contradicting the disjointness of V and W. Hence, $A = f^{-1}(V)$ is an open set in X. Similarly, we see that $B = f^{-1}(W)$ is open in X. Hence X is not connected because $X = A \cup B$ where A and B are nonvoid, disjoint open subsets of X.

Theorem 5 of [1] states that if X and Y are regular spaces then $f\colon X \to Y$ is closure continuous on X if and only if f is continuous on X. A result of Rowe [2] states that a function $f\colon \mathbb{R} \to \mathbb{R}$ is continuous if and only if f satisfies both of the following conditions: (1) f is connected; (2) $f^{-1}(y)$ is closed for each real number y. It is well known that for real-valued functions of a real variable the condition of being connected is equivalent to the Intermediate Value Property (Darboux Property): Given a number C between $f(a)$ and $f(b)$, there exists c between a and b such that $f(c) = C$. It is also well known that if $f\colon \mathbb{R} \to \mathbb{R}$ is such that $f'(x)$ exists for each $x \in \mathbb{R}$ then $f'(x)$ satisfies the Darboux Property. Now, let $f\colon \mathbb{R} \to \mathbb{R}$ be any function such that $f'(x)$ exists but is *discontinuous* on $\mathbb{R}$. Then $f'(x)$ is a *connected function* that is *not closure continuous* on $\mathbb{R}$. (If $f'(x)$ were closure

continuous on $\mathbb{R}$, then $f'(x)$ would be continuous by Theorem 5 of [**1**].) Hence, the converse of Theorem 1 of the present note is false in general.

A pair of conditions each strictly weaker than continuity is called a *factorization* of continuity if together the pair of conditions implies continuity. Thus, for real-valued functions of a real variable, {Darboux Property, closed point inverses} is a factorization of continuity. The example 1 in [**1**] shows that {closure continuity, closed point inverses} fails to be a factorization of continuity. We close this note with the following question: What are some factorizations of continuity with closure continuity as a factor?

REFERENCES

1. D. R. Andrew and Elaine Kirley Whittlesy, Closure continuity, this MONTHLY, 73 (1966) 758–759.
2. C. H. Rowe, Note on a pair of properties which characterize functions, *Bulletin of the American Mathematical Society*, 32 (1926) 285–287.
3. J. Chew, On a converse of the intermediate value theorem, *Journal of Undergraduate Mathematics*, 2 (1972) 79–82.

EDITED BY JOAN P. HUTCHINSON AND STAN WAGON

Dimensional Analysis and Calculus Identities

J. R. TORCZYNSKI
Fluid and Thermal Sciences Department, Sandia National Laboratories, Albuquerque, NM 87185

Introduction. In everyday life, we encounter a variety of units. For example, we might put 10 *gallons* of gasoline into the tank of our 2000-*pound* automobile and then proceed to use it up by driving 200 *miles* at a speed of 50 *miles per hour* during the next 4 *hours*. These common units are measures of certain physical attributes—an hour is a unit of time, a mile is a unit of distance, and so forth. However, we could have just as easily expressed the trip distance d in kilometers. There is nothing special about miles or kilometers: they both express the physical property of "length." Similarly, hours, minutes, and seconds are measures of the physical attribute called "time."

If we denote the trip time in the above example by t, the units of t are time, or $[t] = T$, where we have followed the convention of denoting the units of a variable by $[\dots]$ and the unit of time by T. Note that here we are making a distinction between common units such as hours and fundamental units, or *dimensions*, such as time. When we say that a variable has dimensions of time, we mean that the fundamental physical property of this variable is time. Now this time may be measured in common units like minutes or hours, but the dimension (property, fundamental unit) is time. Similarly, we denote the dimensions of d by $[d] = L$, where the unit of length is denoted by L.

Denoting the average velocity by v, we might be tempted to make the statement that $[v] = V$, where V is the fundamental unit of velocity. This statement is correct; nevertheless, it is more illuminating to say $[v] = L/T$ since velocity has the dimensions of length divided by time. Length, time, and velocity are not three "independent" dimensions, so we would not wish to include all in a list of "fundamental" dimensions. Moreover, in a problem involving only v and the sound speed c, we would include only $V = L/T$, rather than both L and T, in a list of independent units since L and T appear only in this combination in all the problem variables and thus are not independent units *in this problem*.

So far, all the physical quantities we have considered have dimensions associated with them. It is possible, however, to construct "dimensionless" quantities. Angles are a good example of this. An angle θ is defined as the ratio of s, the length of circular arc it intercepts, to r, the length of the radius of the circle. It is clear that $[s] = [r] = L$. We can solve for the dimension of θ as follows:

$$[\theta] = [s/r] = [s]/[r] = L/L = 1.$$

Thus, the angle θ is a dimensionless or "pure" number.

The Buckingham Pi Theorem. Proofs employing dimensional reasoning frequently involve and enhance our understanding of the fundamental issues in mathematical and physical problems [**1**], [**2**]. This method, as applied to physical

systems, was first rigorously described by Buckingham [3], whose Buckingham Pi Theorem, outlined below, expresses the basic principle of dimensional analysis. Given a problem described by N parameters x_i with M' units u_j, then there are $K = N - M$ dimensionless parameters Π_k which fully describe the problem, where $M \leqslant M'$ is the number of independent units. By fully describing the problem, it is meant that the equations of the problem can be recast into a form where only the dimensionless parameters and dimensionless variables appear. Thus, two problems with different values of the x_i are actually the same when made dimensionless if the values of the Π_k are the same. Also, for each function q in the problem relating the x_i such that $q(x_1, \ldots, x_N) = 0$, there is a function p relating the Π_k such that $p(\Pi_1, \ldots, \Pi_K) = 0$. In the case where one such function exists, this result is typically expressed as $\Pi_L = p_L(\Pi_1, \ldots, \Pi_{L-1}, \Pi_{L+1}, \ldots, \Pi_K)$, where tacit use has been made of the implicit function theorem. Except for pathological cases, this inversion can be performed, and we shall assume this to be true for all later examples.

The justification of this theorem is seen in the construction of the dimensionless Π_k. We can write the dimensions of each of the x_i in the following way.

$$[x_i] = \prod_{j=1}^{M'} u_j^{\beta_{ji}}.$$

Choose the following form for each of the Π_k.

$$\Pi_k = \prod_{i=1}^{N} x_i^{\alpha_{ik}}.$$

The dimensions of the Π_k are then as given below.

$$[\Pi_k] = \prod_{i=1}^{N} [x_i]^{\alpha_{ik}}.$$

These equations are combined with the relations for $[x_i]$ to yield explicit expressions for the dimensions of the Π_k.

$$\begin{aligned}[\Pi_k] &= \prod_{i=1}^{N} \prod_{j=1}^{M'} u_j^{\beta_{ji}\alpha_{ik}} \\ &= \prod_{j=1}^{M'} u_j^{(\sum_{i=1}^{N} \beta_{ji}\alpha_{ik})}.\end{aligned}$$

Now we wish the Π_k to be dimensionless, or $[\Pi_k] = 1$ for all k. Hence, the exponents of all the units must vanish. This results in the following matrix equation.

$$\sum_{i=1}^{N} \beta_{ji}\alpha_{ik} = 0.$$

Note that columns of the α-matrix are just the eigenvectors of the β-matrix having an eigenvalue of zero. Let the β-matrix have rank $M \leqslant M'$. Thus, by definition there are M independent units. Since the β-matrix has M linearly independent rows and N columns, it has $K = N - M$ such eigenvectors. Hence, the α-matrix has K columns, corresponding to the K different Π_k. Note that the specific choice of the

Π_k is not unique. For example, $\Delta_k = \Pi_1\Pi_k$ is an equally good choice, corresponding to the new eigenvectors $\alpha_{ik}^{(\Delta)} = \alpha_{1k}^{(\Pi)} + \alpha_{ik}^{(\Pi)}$.

Dimensional Analysis in Physical Science. A few illustrative examples are useful to show how dimensional-analysis ideas may be applied to problems in areas of physical science.

Suppose we wish to find an expression for the angular frequency Ω of a simple pendulum. We take the pendulum bob to be a point mass m suspended from a frictionless pivot by a massless rod of length l in a gravitational field of acceleration g. The maximum angular displacement from the vertical is denoted by θ. We have five parameters: $x_1 = \Omega$, $x_2 = m$, $x_3 = l$, $x_4 = g$, and $x_5 = \theta$. Appearing in this problem are three units: length $u_1 = L$, time $u_2 = T$, and mass $u_3 = M$. The dimensions of the five parameters are given below in terms of these units: $[\Omega] = T^{-1}$, $[m] = M^1$, $[l] = L^1$, $[g] = L^1T^{-2}$, and $[\theta] = 1$. Thus, in this case we have five parameters and three independent units, so we expect $5 - 3 = 2$ dimensionless parameters: $\Pi_1 = \Omega(l/g)^{1/2}$ and $\Pi_2 = \theta$ form one possible set of dimensionless parameters. We presume that there is a function relating the angular frequency Ω to the other four parameters, so there exists a function p such that $p(\Pi_1, \Pi_2) = 0$. Assuming this relation may be solved for Π_1, we find $\Pi_1 = f(\theta)$ or

$$\Omega = (g/l)^{1/2} f(\theta). \tag{1}$$

Here f is some unspecified function.

At this point a few observations may be made. Note that the mass m of the pendulum does not enter into this relationship, so the frequency of a simple pendulum is independent of its mass, a point not immediately obvious. One might argue then that we shouldn't have counted the mass as a parameter, which reduces the number of parameters to four. If this is done, however, the number of units is reduced from three to two (no other parameter has units of mass), so we would still expect $4 - 2 = 2$ dimensionless parameters.

We can also apply physical intuition to (1). If the idea of a pendulum as a time-measuring instrument has any validity, then $f(\theta)$ should have a nonzero limit as θ approaches zero. [In fact, $f(0) = 1$, which exemplifies the fact that in physical problems, dimensionless parameters are often $O(1)$.] Thus, had we neglected to include θ, the finite amplitude of the swing, in our parameter list, we would have found that $p(\Pi_1) = 0$. Hence, Π_1 would be a constant, so $\Omega = C(g/l)^{1/2}$, where C is a constant, dimensionless number. A guess that the constant C is $O(1)$ would be "physically" excellent. Reasoning along similar lines (an exercise for the reader), one is able to show that $f(\theta)$ is an even function of θ.

Let us consider another physical problem (much less laboriously than above). Find the force F on a sphere of diameter D moving with velocity U through an incompressible fluid of density ρ and viscosity μ. The physical quantities have the following units: $[F] = M^1L^1T^{-2}$, $[D] = L^1$, $[U] = L^1T^{-1}$, $[\rho] = M^1L^{-3}$, and $[\mu] = M^1L^{-1}T^{-1}$. With five parameters and three units, we expect two dimensionless parameters, which we choose to be $\Pi_1 = F/\rho U^2D^2$ and $\Pi_2 = \rho UD/\mu$. Thus, the drag force is given by

$$F = \rho U^2 D^2 f(\rho UD/\mu), \tag{2}$$

where again f is an undetermined function. The parameter Π_2 is referred to as the

Reynolds number [4], a dimensionless number of great importance in many problems involving fluid motion.

Physical knowledge can be used as above to say a little more about the function f. It is expected from the equations of motion that the fluid density becomes unimportant (that is, can be left out of our list of parameters) for small values of the Reynolds number, whereas the viscosity becomes unimportant for large values of the Reynolds number. It is left as an exercise for the reader to use these facts to show from (2) or directly that $F = C_1 UD\mu$ for small Reynolds numbers (Stokes' law), and that $F = C_2 \rho U^2 D^2$ for large Reynolds numbers (turbulent drag). Here, C_1 and C_2 are dimensionless constants, coincidentally both $O(1)$.

Dimensional Analysis and Geometry. It is not widely appreciated that dimensional analysis alone is sufficient to prove some elementary *mathematical* theorems.

One of the best examples is the elegant proof of the Pythagorean Theorem [5] employing dimensional analysis. The Pythagorean Theorem states that $a^2 + b^2 = c^2$, where a and b are the legs and c is the hypotenuse of a right triangle. Denote the two acute angles by α and β, as in FIG. 1. Now giving two angles and one side determines a triangle, and by hypothesis one of the angles here is a right angle, so specifying α and c determines all quantities associated with a right triangle. By this statement we mean that there are functions of α and c that describe a, b, β, the area A, and any other quantity related to the triangle. The angles α and β are dimensionless, $[\alpha] = [\beta] = 1$, but a, b, and c have units of length, $[a] = [b] = [c] = L^1$. Thus, from dimensional reasoning, we have the following relations: $a(c, \alpha) = cf_a(\alpha)$, $b(c, \alpha) = cf_b(\alpha)$, and $\beta(c, \alpha) = g(\alpha)$. The area A has the units of length squared, $[A] = L^2$, so we find

$$A(c, \alpha) = c^2 f(\alpha). \tag{3}$$

Here, f, f_a, f_b, and g are unspecified functions (although we, with our understanding of trigonometry, know what these functions are). As an aside, note that values of α and β may be interchanged (either α or β could be called "α"), so $\alpha = g(\beta)$ and hence the function g is its own inverse function, $g = g^{-1}$.

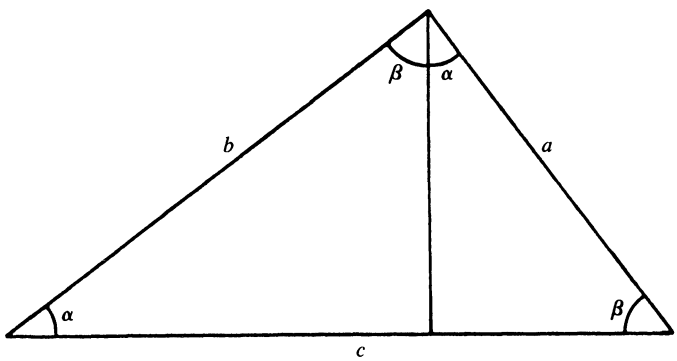


FIG. 1. An arbitrary right triangle.

Partition the right triangle into two right triangles using the altitude from the right-angle vertex to the hypotenuse. One triangle shares acute angle α with the

original triangle; hence, its other acute angle is $\beta = g(\alpha)$. The other right triangle shares acute angle β with the original triangle and therefore has $\alpha = g(\beta)$ as its other acute angle. Note in passing that we have just used dimensional analysis (apparently in a way previously unrecognized) to prove the complementarity of α and β (their sum is a right angle, as shown in FIG. 1): $\alpha + \beta = \theta_R$, or $g(\alpha) = \theta_R - \alpha$, where here we define a right angle to have a measure of θ_R. The hypotenuses of these two new right triangles are a and b, respectively. Clearly, the sum of their areas equals the area of the original triangle

$$A(a,\alpha) + A(b,\alpha) = A(c,\alpha). \tag{4}$$

Next, we use (3), the explicit formula for the area, in (4) and divide out the common factor $f(\alpha)$ to obtain (5), the Pythagorean Theorem.

$$a^2 f(\alpha) + b^2 f(\alpha) = c^2 f(\alpha)$$

$$a^2 + b^2 = c^2. \tag{5}$$

As an aside, suppose there were an additional independent length scale R, say the local curvature of space. For example, we might be interested in spherical trigonometry. Now all triangle quantities must be functions of α, c, and R, so $\beta = g(\alpha, c/R)$ since c/R is a new dimensionless parameter. Although it is still true that $\alpha = g(\beta, c/R)$, our subdivision of the triangle runs into trouble since $\beta \neq g(\alpha, a/R)$ and $\alpha \neq g(\beta, b/R)$. In such a situation, there generally will not be a "Pythagorean Theorem."

In a similar vein, it is possible to prove that the three angles of an arbitrary triangle are supplementary. This proof, although qualitatively the same as the complementarity proof above, is lengthier, so we leave it as an exercise for the interested reader. Hint: draw the segments joining the midpoints of the three sides of the triangle.

Dimensional Analysis and Calculus. It is also possible to use dimensional analysis *alone* to prove some simple calculus identities. More specifically, we examine the relation

$$\frac{d^2x}{dy^2} = -\left(\frac{dy}{dx}\right)^{-3}\frac{d^2y}{dx^2} \tag{6}$$

and prove it using dimensional reasoning. When compared to the conventional calculus proof of (6), the dimensional argument is elegant and straightforward, involving an understanding of the mathematical and physical basis of (6), rather than emphasizing the mechanical aspects of repeated application of the chain rule.

We are interested in finding a formula relating the nth derivative of x from $x = x(y)$ to the derivatives of y from $y = y(x)$. The conventional calculus approach to this problem is straightforward, emphasizing a mechanical use of the chain rule. First, we identify

$$\frac{d}{dy} = \left(\frac{dy}{dx}\right)^{-1}\frac{d}{dx}. \tag{7}$$

Then the desired relation becomes the following:

$$\frac{d^n x}{dy^n} = \left\{ \left(\frac{dy}{dx} \right)^{-1} \frac{d}{dx} \right\}^n x. \tag{8}$$

Let us examine the $n = 2$ case. All one need do is use (8) and painlessly "turn the crank."

$$\begin{aligned} \frac{d^2 x}{dy^2} &= \left\{ \left(\frac{dy}{dx} \right)^{-1} \frac{d}{dx} \right\}^2 x \\ &= \left(\frac{dy}{dx} \right)^{-1} \frac{d}{dx} \left(\frac{dy}{dx} \right)^{-1} \frac{d}{dx} x \\ &= \left(\frac{dy}{dx} \right)^{-1} \frac{d}{dx} \left(\frac{dy}{dx} \right)^{-1} \\ &= \left(\frac{dy}{dx} \right)^{-1} (-1) \left(\frac{dy}{dx} \right)^{-2} \frac{d^2 y}{dx^2} \\ &= -\left(\frac{dy}{dx} \right)^{-3} \frac{d^2 y}{dx^2} \end{aligned}$$

Little insight goes into or comes out of this mechanical type of a proof, and there are ample opportunities for algebra errors.

We now wish to use dimensional analysis to find the explicit formula for the nth derivative in the reversed coordinates, where

$$\frac{d^n x}{dy^n} = \left\{ \left(\frac{dy}{dx} \right)^{-1} \frac{d}{dx} \right\}^n x. \tag{9}$$

Several facts become clear upon reflection. First, the explicit formula depends only on integer powers of the first through nth derivatives. That the formula cannot depend on the coordinates x and y themselves is understood from the fact that translating the function in the x- and y-directions does not affect slopes, curvatures, etc.

$$\frac{d^n x}{dy^n} = F_n\left(\frac{dy}{dx}, \ldots, \frac{d^n y}{dx^n} \right). \tag{10}$$

Second, the explicit formula must be dimensionally correct in both x and y. Third, the explicit formula cannot depend on the coordinate label. In other words, the labels x and y may be exchanged in the explicit formula, again without altering the functional form, so F_n must be the same function in (10) and (11).

$$\frac{d^n y}{dx^n} = F_n\left(\frac{dx}{dy}, \ldots, \frac{d^n x}{dy^n} \right). \tag{11}$$

Let us examine in detail the application of this dimensional reasoning to the proof of the $n = 2$ case. The second derivative is a function of only the first and the second derivatives.

$$\frac{d^2x}{dy^2} = F_2\left(\frac{dy}{dx}, \frac{d^2y}{dx^2}\right). \tag{12}$$

Next, we invoke dimensional correctness in x and y, essentially using the Buckingham Pi Theorem [3], where we take $[x] = X^1$ and $[y] = Y^1$. The left hand side of (12) has one power of x in the numerator and two powers of y in the denominator. Thus, its dimensions are

$$\left[\frac{d^2x}{dy^2}\right] = X^1Y^{-2}. \tag{13}$$

The arguments of F_2 have the following dimensions:

$$\left[\frac{dy}{dx}\right] = X^{-1}Y^1, \tag{14}$$

$$\left[\frac{d^2y}{dx^2}\right] = X^{-2}Y^1. \tag{15}$$

There are three parameters and two independent units, so we expect that there will be $3 - 2 = 1$ dimensionless parameters. There is only one way to combine these quantities to give the dimensionless variable Π.

$$\Pi = \frac{d^2x}{dy^2}\left(\frac{dy}{dx}\right)^3\left(\frac{d^2y}{dx^2}\right)^{-1}. \tag{16}$$

Since by assumption there is a function (F_2) relating the three parameters, we know there is a function $p(\Pi)$ such that $p(\Pi) = 0$. Hence, Π must be a constant, independent of whatever specific function $y = y(x)$ we are considering. We can manipulate (16) into the following form.

$$\frac{d^2x}{dy^2} = \Pi\left(\frac{dy}{dx}\right)^{-3}\frac{d^2y}{dx^2}. \tag{17}$$

Finally, we use the fact that we may interchange x and y.

$$\frac{d^2y}{dx^2} = \Pi\left(\frac{dx}{dy}\right)^{-3}\frac{d^2x}{dy^2}. \tag{18}$$

Comparing the above two equations shows that $\Pi^2 = 1$, so $\Pi = 1$ or -1. One can trivially observe (by thinking about signs of slopes and concavities) that $\Pi = -1$. This yields the desired result.

$$\frac{d^2x}{dy^2} = -\left(\frac{dy}{dx}\right)^{-3}\frac{d^2y}{dx^2}. \tag{19}$$

Note that in this derivation we have not used the chain rule or taken derivatives. Rather, this identity follows from purely dimensional considerations.

An alternate and perhaps more subtle dimensional proof of (19) involves the physical idea of the radius of curvature R_c of a function $y = y(x)$ (see FIG. 2). Since R_c is determined by matching the first and second derivatives of a function to those of a circle,

$$R_c = G\left(\frac{dy}{dx}, \frac{d^2y}{dx^2}\right). \tag{20}$$

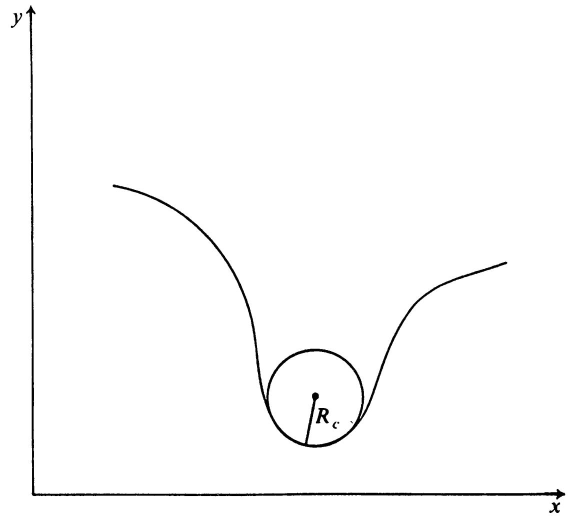


FIG. 2. The radius of curvature.

Since R_c, x, and y all have the dimensions of length,

$$R_c = \left(\frac{d^2y}{dx^2}\right)^{-1} H\left(\frac{dy}{dx}\right). \tag{21}$$

But R_c cannot depend on the coordinate system, so

$$R_c = \left(\frac{d^2x}{dy^2}\right)^{-1} H\left(\frac{dx}{dy}\right). \tag{22}$$

Combining (21) and (22) gives the following:

$$\frac{d^2x}{dy^2} = \frac{d^2y}{dx^2} K\left(\frac{dy}{dx}\right). \tag{23}$$

We are now essentially at (12) and can use the same line of reasoning to complete the proof. Thus, from our physical understanding of the radius of curvature, the *mathematical* identity is proved even more directly and elegantly.

One may inquire whether the derivative identities may be proved by dimensional analysis for $n > 2$. The answer is yes, almost. The functional form is always

correctly given, but some of the dimensionless constants remain undetermined. As an example, we may proceed as above for the $n = 3$ case, using the results already obtained for smaller n. In this case we find

$$\frac{d^3x}{dy^3} = -\left(\frac{dy}{dx}\right)^{-4}\frac{d^3y}{dx^3} + C\left(\frac{dy}{dx}\right)^{-5}\left(\frac{d^2y}{dx^2}\right)^2,$$

where C is an undetermined constant. Dimensional analysis takes us no further. To determine C, we would have to use a test function (such as $y = x^{-1}$ or $y = e^x$), which yields $C = 3$. For larger n, there are progressively more undetermined constants, and the method becomes impractically cumbersome.

Conclusions. It is seen that dimensional reasoning alone is sufficient to prove the identity

$$\frac{d^2x}{dy^2} = -\left(\frac{dy}{dx}\right)^{-3}\frac{d^2y}{dx^2}.$$

This proof does not depend upon taking derivatives. Instead, it stresses insight into the underlying physical basis of the identity (involving the radius of curvature). Such a method of proof is seen to be elegant and is useful for developing physical and mathematical intuition rather than mechanical proficiency.

Acknowledgments. The author wishes to acknowledge the pleasure and utility of many discussions with Professor Hans W. Liepmann, one of the undisputed masters of dimensional analysis.

REFERENCES

1. Lord Rayleigh, *Nature*, 95 (March, 1915) 66.
2. L. I. Sedov, Similarity and Dimensional Analysis in Mechanics, Academic Press, New York, 1959.
3. E. Buckingham, *Phys. Rev.*, 4(4) (1914) 345–376.
4. O. Reynolds, Papers on Mathematical and Physical Subjects, Cambridge University Press, 1903.
5. G. I. Barenblatt, Similarity, Self-Similarity, and Intermediate Asymptotics, Consultants Bureau (Plenum Press), New York, 1979, p. 23.

Reconsidering Area Approximations

GEORGE P. RICHARDSON
The Rockefeller College of Public Affairs and Policy, State University of New York at Albany, Albany, NY 12222

A recent survey of calculus texts for adoption revealed that all contained presentations of the familiar method of approximating a definite integral by summing the areas of n trapezoids of equal width h,

$$A \doteq \sum_{i=1}^{n} \frac{f(x_{i-1}) + f(x_i)}{2} h.$$

Only one text contained any mention of the simpler approximation obtained by summing the areas of rectangles computed at midpoints,

$$A \doteq \sum_{i=1}^{n} f\left[\frac{x_{i-1} + x_i}{2}\right] h,$$

and there the midpoint-rectangle method was quickly dismissed in an example. The overwhelming tendency to ignore the midpoint-rectangle method in modern texts is unfortunate, and somewhat puzzling, because it is conceptually simpler than the trapezoid approximation and is actually more accurate. A good case can be made for the claim that the midpoint-rectangle method is the first numerical approximation to the definite integral that students should see.

Although the greater accuracy of the midpoint-rectangle approximation is well known, it is nonetheless surprising to many, perhaps partly because we are so used to seeing the trapezoid method emphasized in introductory texts. Yet, the result is easy to see geometrically (see FIGURE 1). The comparison of the area approximations is facilitated by noting that the area of the midpoint-rectangle is equal to the area of another trapezoid determined by the tangent to the graph of f at the midpoint. From the figure it is then evident that as long as $f''(x)$ is continuous and nonzero over $[x_{i-1}, x_i]$, so that the direction of curvature does not change, the area representing the midpoint-rectangle error will always be less than the area representing the trapezoid error.

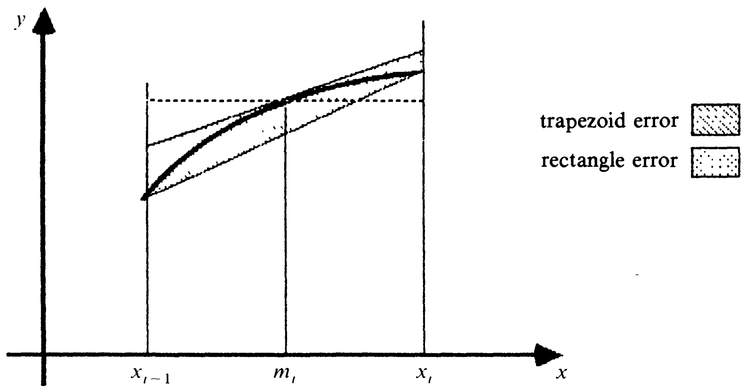


FIG. 1. Geometric comparison of the midpoint-rectangle and trapeziod approximations for $\int_{x_{i-1}}^{x_i} f(x)dx$.

Most calculus texts state without proof a theorem equivalent to the following. If f has a continuous second derivative on $[a, b]$ then the error e_T in approximating

$$\int_b^a f(x)\,dx$$

using n trapezoids of width $h = (b - a)/n$ satisfies

$$0 \leqslant e_T \leqslant \frac{1}{12}h^2(b-a)\Big[\max_{a \leqslant x \leqslant b}|f''(x)|\Big].$$

The corresponding error bound e_R for n rectangles, which is not commonly mentioned in calculus texts, is

$$0 \leqslant e_R \leqslant \frac{1}{24}h^2(b-a)\Big[\max_{a \leqslant x \leqslant b}|f''(x)|\Big].$$

Proofs can be found in most numerical analysis texts, e.g. [**1**], [**2**], [**3**], [**6**, pp. 361–400]. While the inequalities here do not support a precise conclusion, the denominators in these error bounds certainly hint that the midpoint-rectangle approximation is likely to be about twice as accurate as the trapezoid method.

The following derivation of series expressions for the approximation errors of the two methods (stimulated by [**4**, pp. 80–84]) is of interest because it is accessible to first-year calculus students, is apparently not well known, and gives more explicit statements of the approximation errors.

Suppose f satisfies the hypotheses of Taylor's Theorem over $[a, b]$. Partition $[a, b]$ into n subintervals $[x_{i-1}, x_i]$ each of length h. Let m_i represent the midpoint of $[x_{i-1}, x_i]$, and for any x in $[x_{i-1}, x_i]$ expand $f(x)$ in a Taylor series about m_i. On integrating over $[x_{i-1}, x_i]$, the even powers in the resulting series cancel, leaving

$$\int_{x_{i-1}}^{x_i} f(x)\,dx = f(m_i)h + \frac{h^3}{2^2 3!}f''(m_i) + \frac{h^5}{2^4 5!}f^{(4)}(m_i) + \cdots. \tag{1}$$

Summing over the n subintervals yields the rectangle approximation and a series expression for its error:

$$\int_a^b f(x)\,dx = \sum_{i=1}^{n} f(m_i)h + \frac{h^3}{2^2 3!}\sum_{i=1}^{n} f''(m_i) + \frac{h^5}{2^4 5!}\sum_{i=1}^{n} f^{(4)}(m_i) + \cdots. \tag{2}$$

To derive the comparable result for the trapezoid rule, use the same Taylor series to express $f(x_{i-1})$ and $f(x_i)$ in terms of $f(m_i)$ and its derivatives. Averaging $f(x_{i-1})$ and $f(x_i)$ and then solving for $f(m_i)$ produces

$$f(m_i) = \frac{f(x_{i-1}) + f(x_i)}{2} - \frac{h^2}{2^2 2!}f''(m_i) - \frac{h^4}{2^4 4!}f^{(4)}(m_i) - \cdots. \tag{3}$$

Substituting the right side of (3) into (1) in place of $f(m_i)$ and collecting like powers of h yields

$$\int_{x_{i-1}}^{x_i} f(x)\,dx = \frac{f(x_{i-1}) + f(x_i)}{2}h - \frac{2h^3}{2^2 3!}f''(m_i) - \frac{4h^5}{2^4 5!}f^{(4)}(m_i) - \cdots,$$

which gives the following result for the trapezoid rule, comparable to (2):

$$\begin{aligned}\int_a^b f(x)\,dx = \sum_{i=1}^{n} \frac{f(x_{i-1}) + f(x_i)}{2}h - \frac{2h^3}{2^2 3!}\sum_{i=1}^{n} f''(m_i) \\ - \frac{4h^5}{2^4 5!}\sum_{i=1}^{n} f^{(4)}(m_i) - \cdots.\end{aligned} \tag{4}$$

From the series error terms in (2) and (4) it is again evident that the rectangle approximation is of the same order as the trapezoid approximation but is likely to be about twice as accurate. More precisely, the error terms show that both methods are exact for linear functions, while the midpoint-rectangle approximation has exactly half the error of the trapezoid approximation for quadratic and cubic polynomials. For functions with nonzero higher derivatives over $[a, b]$, the smaller error inherent in the midpoint-rectangle approximation will theoretically become evident once h is sufficiently small.

There is both puzzle and promise in these results. The puzzle is that the midpoint-rectangle method is almost uniformly ignored in introductory calculus texts, and the trapezoid rule universally emphasized, although the midpoint rule is likely to be more accurate. Moreover, the midpoint-rectangle method is conceptually simpler and is closer to the fundamental notion of a Riemann sum. For building intuition about the definite integral, probably every student should experiment with a simple computer implementation of the method.

Besides their pedagogical implications, the series expansions for the approximation errors in (2) and (4) provide a further benefit. As suggested by FIGURE 1, the errors have opposite signs. Taking advantage of that fact, one can easily note that doubling (2) and adding it to (4) eliminates the terms involving h^3 [4, pp. 80–84]. Thus on dividing by 3 we obtain an approximation of the definite integral that is accurate to an error proportional to h^5:

$$\int_a^b f(x)\,dx = \frac{1}{6}\sum_{i=1}^{n}\left[f(x_{i-1}) + 4f(m_i) + f(x_i)\right]h - \frac{h^5}{2^3 3\cdot 5!}\sum_{i=1}^{n} f^{(4)}(m_i)$$

$$- \frac{2h^7}{2^5 3\cdot 7!}\sum_{i=1}^{n} f^{(6)}(m_i) - \cdots .$$

This substantial improvement is, of course, Simpson's rule, and equations (2) and (4) provide an elegant derivation. (One introductory calculus-with-applications text has a nicely intuitive development along these lines without the detailed error terms [3, pp. 454–457].) The same technique done with trapezoid approximations with n and $2n$ trapezoids also yields Simpson's rule and can be efficiently iterated in Romberg's method [5] to produce much more accurate, higher-order approximations.

REFERENCES

1. Ake Bjorck and Germund Dahlquist, Numerical Methods, Prentice-Hall, Englewood Cliffs, N.J., 1974.
2. Richard L. Burden and J. Douglas Faires, Numerical Analysis, 3rd ed., Prindle, Weber & Schmidt, Boston, 1985.
3. Larry J. Goldstein, David C. Lay, David I. Schneider, Calculus and its Applications, 4th ed., Prentice-Hall, 1987.
4. Robert Sedgewick, Algorithms, Addison-Wesley, 1984.
5. Stanley Wagon, Evaluating Definite Integrals on a Computer: Theory and Practice, UMAP unit 432, EDC/UMAP, Newton, Mass., 1980.
6. David M. Young and Robert Todd Gregory, A Survey of Numerical Mathematics, vol. 1, Addison-Wesley, Reading, Mass., 1972.

EDITORS' NOTE: We have received several letters concerning F. Brauer's article [1], which presented a proof of a version of Taylor's theorem. First, James Smith (San Francisco State Univ.) pointed out that the same approach appeared in a 1950 MONTHLY article by C. L. Seebeck [2]. Moreover, Seebeck obtained the Lagrange form of the remainder. Sherman Stein (Univ. of Calif., Davis) observed that essentially this approach appeared in the second edition of his calculus text [3]. However, the third and fourth editions gave Taylor's theorem with the integral form of the remainder, to reinforce integration by parts and the Fundamental Theorem of Calculus, and because the integral form is stronger than the Lagrange form. Stein adds, "Although the formulas that appear in the statement and proof of the integral version are messier than in the Lagrange version, I believe that the concepts are simpler, more natural, and more important." Finally, Peter Fowler (Calif. State Univ., Hayward) and Rudolf Výborný (Univ. of Queensland, Australia) pointed out that Brauer's version should have explicitly stated the assumption of the continuity of the $(n+1)$st derivative of f at x. For more on this assumption see the following article by S. D. Chatterji.

REFERENCES

1. F. Brauer, A simplification of Taylor's theorem, this MONTHLY, 94 (1987) 453–455.
2. L. Seebeck, Jr., A note on Taylor's theorem, this MONTHLY, 57 (1950) 32–34 (reprinted in Selected Papers on Calculus, ed. by T. Apostol, Math. Assoc. Amer., Washington, 1969).
3. S. K. Stein, Calculus and Analytic Geometry, second ed., McGraw-Hill, New York, 1977.

A Frequent Oversight Concerning the Integrability of Derivatives

S. D. CHATTERJI
Dépt. de Math., Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

In a recent article in this MONTHLY [1], F. Brauer proposed an elementary presentation of the following form of Taylor's theorem: "Suppose that for some integer $n \geqslant 0$ the functions $f, f', \dots, f^{(n)}$ are continuous on $[a, b]$ and that $f^{(n+1)}$ exists and satisfies $|f^{(n+1)}(t)| \leqslant M$ on (a, b). Then on $[a, b]$,

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + R_n(x)$$

with

$$|R_n(x)| \leqslant \frac{M}{(n+1)!}|x-a|^{n+1}."$$

The proof is based on the fact that if $g\colon [a, b] \to \mathbb{R}$ is differentiable and g' is bounded on $[a, b]$ then

$$\int_a^b g'(x)\,dx = g(b) - g(a). \tag{1}$$

However, the validity of (1) cannot be considered to be an entirely standard elementary fact if g' is not assumed to be continuous or at least Riemann integrable on $[a, b]$. The possibility that g' need not be Riemann integrable seems to have been overlooked.

PROBLEMS AND SOLUTIONS

EDITED BY PAUL T. BATEMAN, HAROLD G. DIAMOND, KENNETH B. STOLARSKY (ADVANCED PROBLEMS), AND DOUGLAS B. WEST (ELEMENTARY PROBLEMS)

*For instructions about submitting **proposed** problems for publication in this department see the inside front cover. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the address given on the inside front cover. Two copies suffice.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

For instructions about submitting solutions of Problems, which should be mailed before February 28, 1989, see the inside front cover. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label if an acknowledgement is desired.

ELEMENTARY PROBLEMS

E 3282. *Proposed by D. M. Milosevic, Pranjani, Yugoslavia.*

Given a triangle ABC, suppose that w_a, w_b, w_c are the lengths of the angle bisectors, s is the semiperimeter, r is the radius of the incircle, and R is the radius of the circumcircle. Prove that

$$w_a^2 + w_b^2 + w_c^2 \leqslant s^2 - r(R/2 - r),$$

with equality if and only if the triangle is equilateral.

E 3283. *Proposed by Orrin Frink*, Pennsylvania State University.*

Every simple closed polygon in the plane has three centroids, namely the centroid of its vertex set, the centroid of its boundary, and the centroid of its interior. In general all three are distinct.

(a) In the case of a triangle show that these three centroids coincide if and only if the triangle is equilateral.

(b) Which of the three centroids are affine invariants?

E 3284. *Proposed by Paul Erdős, Hungarian Academy of Sciences, Budapest.*

Determine the minimum number of edges in a triangle-free graph on $2n$ vertices whose complement contains no complete graph on n vertices.

*Professor Frink died on March 4, 1988.

E 3285. *Proposed by W. K. Nicholson, University of Calgary.*

Let R be a simple ring with unity element, i.e., a ring with unity element having no two-sided ideals other than itself and $\{0\}$. Let G be a multiplicative subgroup of the set of invertible elements of R.

(a) Under the assumption that either $x \in G$ or $1 - x \in G$ for every $x \in R$, show that R is a division ring.

(b) If, in addition to the assumption of (a), we assume that $-1 \in G$, show that $G = R \setminus \{0\}$.

E 3286. *Proposed by William C. Waterhouse, Pennsylvania State University.*

Suppose A, B, C, D are n by n matrices and C', D' denote the transposes of C, D. In a comment on problem 6057 [1987, 1020], M. J. Pelling remarked that the condition

$$CD' + DC' = 0 \tag{$*$}$$

implies

$$\det\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det(AD' + BC') \tag{$**$}$$

if D is nonsingular, but gave an example with singular D satisfying $(*)$ in which one side of $(**)$ is 1 and the other side is -1. Prove that in any case $(*)$ implies

$$\det\begin{pmatrix} A & B \\ C & D \end{pmatrix}^2 = \det(AD' + BC')^2.$$

E 3287. *Proposed by David M. Bloom, Brooklyn College.*

It is easy to see that for each x in $(1, e)$ there is a unique $y = g(x)$ in (e, ∞) such that $x^y = y^x$, or equivalently $(\log y)/y = (\log x)/x$. For example, $g(2) = 4$. Prove that if $h(x) = x^{g(x)}$, then h is a decreasing function on $(1, e)$.

E 3288. *Proposed by John J. Cade, Pikeville College, Pikeville, KY.*

It is well known that if a, m, and n are positive integers with $a > 1$, then

$$(a^m - 1, a^n - 1) = a^{(m,n)} - 1,$$

where (m, n) is the greatest common divisor of m and n. Determine $(a^m - 1, a^n + 1)$ and $(a^m + 1, a^n + 1)$.

SOLUTIONS OF ELEMENTARY PROBLEMS

Angle Bisectors

E 2966 [1982, 594]. *Proposed by P. J. Giblin, University of Liverpool, U.K.*

A, B, P_1, P_2, P_3 are distinct points in the plane. P_iP_jA, P_iP_jB are proper triangles, i.e., no two of P_1, P_2, P_3 are collinear with A or with B. The anticlockwise angles from AP_1 to AP_2, AP_1 to AP_3, BP_1 to BP_2, BP_1 to BP_3 are $\theta_1, \theta_2, \phi_1, \phi_2$. If

$a_i = AP_i$, $b_i = BP_i$ and if the relations

$$\frac{\sin\theta_1}{\sin\phi_1}\frac{a_3}{b_3} = \frac{\sin\theta_2}{\sin\phi_2}\frac{a_2}{b_2} = \frac{\sin(\theta_2-\theta_1)}{\sin(\phi_2-\phi_1)}\frac{a_1}{b_1}$$

hold, show that the angle AP_iB has the same pair of bisectors as one of the angles of the triangle $P_1P_2P_3$. (Possibly the internal bisector of one angle is the external bisector of the other.)

No complete solution was received. A solution is sketched on pp. 182–183 of the paper, "Source genericity of caustics in the plane," by J. W. Bruce, P. J. Giblin, and C. G. Gibson, *Quart. J. Math. Oxford* (2) 33 (1982) 169–190.

L. Kuipers (Switzerland) solved the altered problem in which the first of the three ratios assumed to be equal is replaced by its reciprocal (but in which the conclusion is unchanged). In the above article Bruce, Giblin, and Gibson obtained not only the assertion of the problem but also its converse; in addition they proved that in fact all three ratios must be ± 1 (and so respectively equal to their reciprocals). Thus Kuipers' argument is consistent with the assertion of the problem but does not establish it as such.

Matching Distances to Vertices

E 3135 [1986, 215]. *Proposed by Hüseyin Demir, Middle East Technical University, Ankara, Turkey.*

For a scalene triangle ABC inscribed in a circle, prove that there is a point D on the circle whose distance from the opposite vertex is the sum of its distances from the other two vertices, and construct D with ruler and compass.

Solution I by J. Leech, University of Stirling, Scotland. Suppose $BC > AC > AB$. Perturb BAC into an isosceles triangle BXY by determining X on the ray BA and Y on BC such that $BX = BY = AC$. The circumcircle of BXY meets the circumcircle of BAC at the required point D. To prove D has the required property, extend AD to a point Z such that $DZ = DC$. Then triangle DZC is similar to triangle BXY, because the angles at D and B are both supplementary to angle ADC. Consequently, triangle DBY is congruent to triangle ZAC, since $AC = BY$ by construction, the angles at A and B are equal and angle BDY = angle BXY = angle DZC. Hence $DB = ZA = DA + DC$.

Solution II by P. Tzermias (student), University of Patras, Greece. Let a, b, c be the lengths of the sides opposite A, B, C, and let x, y, z be the lengths of DA, DB, DC. We seek point D such that $y = x + z$. By Ptolemy's Theorem, $ax + cz = by$. Substituting for y yields $x/z = (c-b)/(b-a)$. Thus, D lies on the "Circle of Apollonius" determined by A and C using the ratio $(c-b)/(b-a)$. This circle has a standard construction (see N. Altshiller-Court, *College Geometry*, 1952, p. 15).

Editorial comments. Since $c - b$ and $b - a$ must have the same sign, D must lie on the arc cut by the side of intermediate length. Consequently, if ABC is isosceles, then D can only be the vertex common to the two equal sides. On the other hand, if ABC is equilateral, then D can be any point on the circumference; Leech's two circles then coincide.

Several solvers noted that the existence of D follows from the Intermediate Value Theorem. If D on the arc opposite B is close to the shortest side of ABC, then its distance to B is less than $DA + DC$, but if D is close to the longest side, then $DB > DA + DC$.

Other solutions independent of Ptolemy's Theorem were submitted by J. Dou (Spain), L. Kuipers (Switzerland), and by P. L. Hon (Hong Kong).

E. Morgantini (Italy) submitted a paper entitled "Una Quartica Bicircolare Della Geometria Del Triangolo" making reference to this problem.

In addition to the solvers mentioned above, correct solutions were received from S. Arslanagić (Yugoslavia), A. Bager, H. Eves, J. Fukuta (Japan), H. Kappus (W. Germany), O. P. Lossers (Netherlands), J. P. Robertson, J. S. Robertson, V. Schindler (E. Germany), R. A. Simon (Chile), B. A. Troesch, M. Vowe (Switzerland), and the proposer.

Large Discs in Convex Unions

E 3139 [1986, 216]. *Proposed by Andrew Vince, University of Florida, Gainesville.*

Let $G_1, \ldots, G_N$ be closed plane convex sets, each of diameter less than or equal to 1 and such that $\bigcup_{i=m}^{n} G_i$ is convex for all (m, n) with $1 \leqslant m \leqslant n \leqslant N$. Prove or disprove: the union of the G_i can contain no disc of diameter greater than $\sqrt{3}$.

Solution by J. Pach, Hungarian Academy of Sciences, Budapest, and C. A. Rogers, University College, London. The assertion of the problem is false; discs of arbitrarily large diameter are possible.

In 1983 we gave the following Proposition (cf. [**1**]):

PROPOSITION. *For every natural number n, there exists a sequence $C_1^{(n)}, C_2^{(n)}, \ldots, C_N^{(n)}$ of compact convex sets satisfying*:

(1) diam $C_r^{(n)} \leqslant 1$ *for* $1 \leqslant r \leqslant N$;

(2) $\bigcup_{i=r}^{s} C_i^{(n)}$ *is convex for* $1 \leqslant r < s \leqslant N$;

and

(3) $\bigcup_{i=1}^{N} C_i^{(n)}$ *contains a disc of radius* n.

We did not prove this result explicitly but asserted that it could be established using ideas similar to those of §2 of our paper. Since §2 is rather complicated, it seems worth giving a fairly simple proof of this proposition.

Consider the four parabolic arcs defined by

$$y = p^{(j)}(x), \qquad -2n \leqslant x \leqslant 2n,$$

for $j = 1, 2, 3, 4$, where

$$\begin{aligned} p^{(1)}(x) &= 4n^2 + 1 - x^2, \\ p^{(2)}(x) &= 4n^2 - x^2, \\ p^{(3)}(x) &= -4n^2 + x^2, \\ p^{(4)}(x) &= -4n^2 - 1 + x^2. \end{aligned}$$

We introduce the sequences

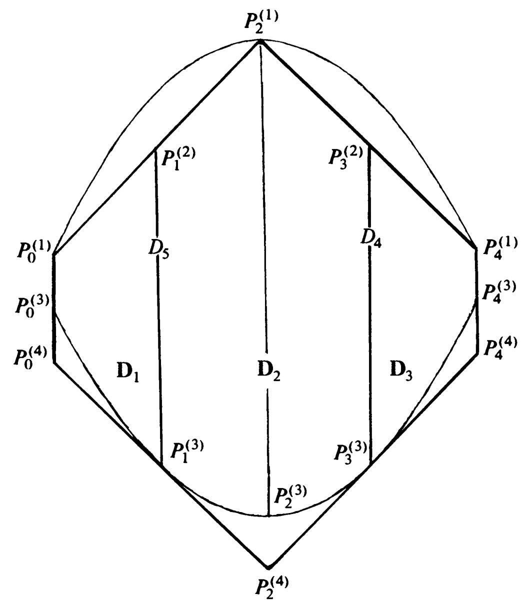
$$P_k^{(j)} = \left(-2n + k, p^{(j)}(-2n + k)\right), \qquad 0 \leqslant k \leqslant 4n,$$

of points on these arcs, $1 \leqslant j \leqslant 4$. We define a sequence $D_1, D_2, \ldots, D_{4n+1}$ of convex domains as follows.

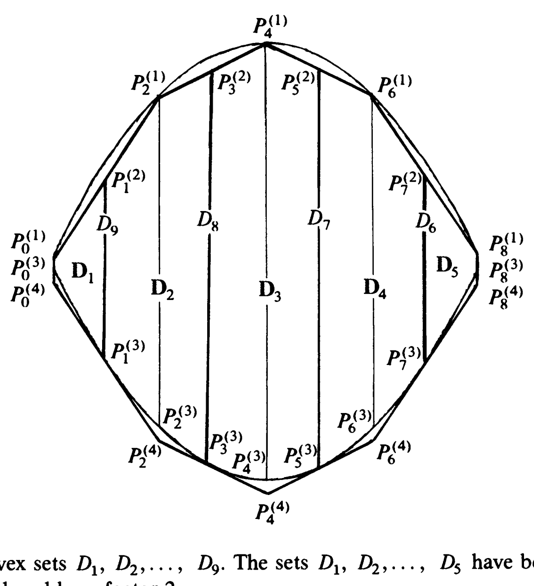
D_1 is the trapezium with vertices $P_0^{(1)}, P_1^{(2)}, P_1^{(3)}, P_0^{(4)}$. For $2 \leqslant i \leqslant 2n$, D_i is the double trapezium $P_{2i-3}^{(2)}, P_{2i-2}^{(1)}, P_{2i-1}^{(2)}, P_{2i-1}^{(3)}, P_{2i-2}^{(4)}, P_{2i-3}^{(3)}$. Finally D_{2n+1} is the trapezium $P_{4n-1}^{(2)}, P_{4n}^{(1)}, P_{4n}^{(4)}, P_{4n-1}^{(3)}$.

For $2n+2 \leqslant i \leqslant 4n+1$, D_i is the region bounded by the line segments $P_{8n+2-2i}^{(3)}, P_{8n+2-2i}^{(1)}$ and $P_{8n+4-2i}^{(3)}, P_{8n+4-2i}^{(1)}$, the arc of the first parabola joining $P_{8n+2-2i}^{(1)}$ and $P_{8n+4-2i}^{(1)}$, and the arc of the third parabola joining $P_{8n+2-2i}^{(3)}$ and $P_{8n+4-2i}^{(3)}$.

See the figures showing these sets in the cases $n = 1$ and $n = 2$.



The case $n = 1$. The convex sets D_1, D_2, D_3, D_4, D_5. The sets D_1, D_2, D_3 have been emphasized. The vertical scale has been decreased by a factor 2.



The case $n = 2$. The convex sets $D_1, D_2, \ldots, D_9$. The sets $D_1, D_2, \ldots, D_5$ have been emphasized. The vertical scale has been reduced by a factor 2.

Note that the sides of the trapezia and double trapezia that are not parallel to the y-axis are tangents to the second parabola at points of the form $P^{(2)}_{2k+1}$ or tangents to the third parabola at points of the form $P^{(4)}_{2k+1}$.

It is easy to verify that the union

$$\bigcup_{i=r}^{s} D_i$$

is convex for $1 \leqslant r < s \leqslant 4n+1$. Define further sets $D_{4n+2}, D_{4n+3}, \ldots, D_{(4n+1)N}$ by taking

$$D_{(4n+1)r+i}, \quad 1 \leqslant i \leqslant 4n+1, \qquad 1 \leqslant r < N,$$

to be the set obtained by translating the set D_i by distance r in the direction of the positive y-axis. It is easy to verify that

$$\bigcup_{i=r}^{s} D_i$$

is convex for $1 \leqslant r < s \leqslant (4n+1)N$. Further $\bigcup_{i=1}^{(4n+1)N} D_i$ contains the rectangle

$$-2n \leqslant x \leqslant 2n, \qquad 0 \leqslant y \leqslant N.$$

Let C_i, $1 \leqslant i \leqslant (4n+1)N$, be obtained from D_i by dividing all the y coordinates by $N/(4n)$. Then each partial union

$$\bigcup_{i=r}^{s} C_i, \qquad 1 \leqslant r < s \leqslant (4n+1)N,$$

is convex, and the union of the $(4n+1)N$ sets contains the square

$$-2n \leqslant x \leqslant 2n, \qquad 0 \leqslant y \leqslant 4n,$$

and so contains a circular disc of radius $2n$. Further, if $\varepsilon > 0$, is given, and N is taken to be sufficiently large, each set C_i has diameter less than $2(1+\varepsilon)$. The required result follows by a change of scale.

1. J. Pach and C. A. Rogers, Partly convex Peano curves, *Bull. London Math. Soc.*, 15 (1983) 321–328.

Editorial Comment. No solutions were received before the customary deadline. However, G. B. Purdy called our attention to the Pach–Rogers paper [1] and suggested that we invite Pach and Rogers to prepare a detailed proof of their proposition. We are grateful to Purdy for this suggestion and to Pach and Rogers for accepting our invitation.

Another Geometric Inequality

E 3146* [1986, 299]. *Proposed by Walther Janous, Ursulinengymnasium, Innsbruck, Austria.*

Prove or disprove the following statement: If a, b, c are the sides of a triangle and $s = (a+b+c)/2$, then

$$2s\{\sqrt{s-a} + \sqrt{s-b} + \sqrt{s-c}\} \leqslant 3\left\{\sqrt{bc(s-a)} + \sqrt{ca(s-b)} + \sqrt{ab(s-c)}\right\}.$$

Solution by Magda P. Nakassis, Gaithersburg, MD. The statement is correct, and the two sides are equal iff $a = b = c$, even when degenerate triangles are considered.

If $s = 0$ there is nothing to prove. If $s > 0$, then there are angles u, v, w between 0 and $\pi/2$ such that $\cos^2 u = (s-a)/s, \cos^2 v = (s-b)/s, \cos^2 w = (s-c)/s$. These satisfy $\sin^2 u = a/s, \sin^2 v = b/s, \sin^2 w = c/s$, and $\cos^2 u + \cos^2 v + \cos^2 w = 1$. Upon division by $s^{3/2}$, the desired inequality now becomes

$$2(\cos u + \cos v + \cos w) \leqslant 3(\cos u \sin v \sin w + \sin u \cos v \sin w + \sin u \sin v \cos w)$$

for three angles between 0 and $\pi/2$ whose squared cosines sum to 1.

We rewrite the right side as $3(\cos u \cos v \cos w - \cos(u+v+w))$. Because the squared cosines sum to 1, we may assume $\cos w \leqslant \sqrt{1/3}$. For a fixed w in this range, $\cos(u+v+w)$ attains a maximum value $f(w)$, which will help put a lower bound on the right side. We compute $f(w)$. Since $0 < u+v+w < 3\pi/2$, for a fixed w the maximum of $\cos(u+v+w)$ occurs at a local extremum of $u+v+w$ or on the boundary.

For the boundary points, one of u or v reaches 0 or $\pi/2$. If either reaches 0, then by the constraint w and the other must equal $\pi/2$, and if either reaches $\pi/2$, then the other must equal $\pi/2 - w$. In either case $u+v+w = \pi$ and $\cos(u+v+w)$ is minimized.

Using a Lagrange multiplier for the constraint $\cos^2 u + \cos^2 v = 1 - \cos^2 w$, an interior extremum of $u+v+w$ must satisfy $1 - 2\lambda \cos u \sin u = 0 = 1 - 2\lambda \cos v \sin v$, or $\sin 2u = \sin 2v$. With $u, v \in [0, \pi/2]$, this implies $u = v$ or $u + v = \pi/2$. If the latter, then w must be $\pi/2$ and $u+v+w = \pi$, which again minimizes $\cos(u+v+w)$. Hence $u = v$ for the only extremum that may have $\cos(u+v+w) > -1$.

We now want the value of $f(w) = \cos(2u+w)$, where $2\cos^2 u = 1 - \cos^2 w$. Since $\cos 2u = 2\cos^2 u - 1$, we can write this in terms of $z = \cos w$ as $f(w) = -z^3 - (1-z^2)\sqrt{1+z^2}$. Note that this value is negative, but that when $0 < z < 1$, it exceeds -1. Hence when $z > 0$ this is the unique point where the maximum is attained.

Having replaced $\cos(u+v+w)$ by $f(w)$, we can substitute $x = \cos u$, $y = \cos v$, $z = \cos w$. Now it suffices to prove that the function $h = 3(xyz + g(z)) - 2(x+y+z)$ is non-negative for each fixed z with $z \leqslant \sqrt{1/3}$, where $g(z) = -f(w) = z^3 + (1-z^2)\sqrt{1+z^2}$, and we have the constraints $x^2 + y^2 + z^2 = 1$ and $x, y, z \geqslant 0$.

Again we use Lagrange multipliers to minimize h over x, y for each fixed value of z. An internal extremum must satisfy $3yz - 2 - 2\lambda x = 0 = 3xz - 2 - 2\lambda y$. This yields $x = y$ or $3z = -2\lambda$. In the latter case, substitution yields $3z(x+y) = 2$. However, since $x^2 + y^2$ is fixed, $x + y$ is maximized when $x = y$, and hence we have $3z(x+y) \leqslant 3z(2\sqrt{(1-z^2)/2})$. This is strictly increasing when $0 \leqslant z \leqslant \sqrt{1/2}$, and it reaches 2 at $z = \sqrt{1/3}$. Hence when $z \leqslant \sqrt{1/3}$ this case lies in the domain only if $x = y = z$, and we may assume $x = y$ for the internal extremum.

When z is fixed, at the boundary points one of x or y is 0 and the other is $\sqrt{1-z^2}$. At such points h reduces to $h(z) = 3z^3 + 3(1-z^2)\sqrt{1+z^2} - 2z - 2\sqrt{1-z^2}$.

Note that $h(0) = 1$ and $h(\sqrt{1/3}) = (3 - 2\sqrt{2})\sqrt{1/3}$, which are positive. In fact, $h(z)$ is strictly decreasing and hence always positive in this range. Differentiating and solving for roots (by squaring both sides) is rather messy; another way to show

positivity of h is via Taylor series expansion (for the square roots) and remainder analysis.

This leaves only the internal extremum $x = y = \sqrt{(1 - z^2)/2}$. Here $h(z) = 3z^3 + 3(1 - z^2)(z/2 + \sqrt{1 + z^2}) - 2z - 4\sqrt{(1 - z^2)/2}$. Now $h(0) = 3 - 2\sqrt{2}$ and $h(\sqrt{1/3}) = 0$. Again h is strictly decreasing for $0 < z < \sqrt{1/3}$, which can be shown by similar means.

Note that $h = 0$ requires $x = y = z = \sqrt{1/3}$, i.e., $a = b = c$.

Also solved by G. Turnwald (West Germany) and I. A. Sakmar (Turkey), and partially solved by P. Khajeh-Khalili. Three incorrect solutions were received.

A Bernoulli Inequality

E 3160 [1986, 565]. *Proposed by L. I. Nicolaescu, University Al. I. Cuza, Iassy, Romania.*

Prove that

$$\frac{(4n)!(4n-2)!}{\{(2n)!\}^4} \leqslant \left|\frac{B_{4n}B_{4n-2}}{B_{2n}^4}\right|, \qquad n \geqslant 2,$$

where B_n denotes the nth Bernoulli number.

Solution I by A. A. Jagers, Technische Hogeschool Twente, Enschede, The Netherlands. Since $|B_{2n}| = 2(2n)!\zeta(2n)/(2\pi)^{2n}$, where ζ denotes the Riemann zeta function $\zeta(2n) = \sum_{k=1}^{\infty} k^{-2n}$, the given inequality is equivalent to $\pi^2\zeta(4n)\zeta(4n-2) \geqslant \zeta(2n)^4$. This inequality is valid for all $n \geqslant 1$, because $\zeta(4n-2) > \zeta(4n) > 1$ and $\zeta(2n) \leqslant \zeta(2) = \pi^2/6 < \sqrt{\pi}$.

Solution II by Robert E. Shafer, Berkeley, CA. Let

$$F(n) = \pi^2\zeta(4n)\zeta(4n-2)/\zeta(2n)^4.$$

The desired inequality is equivalent to $F(n) \geqslant 1$ for integers $n \geqslant 2$. We prove more strongly that $F(n+1) > F(n)$ for $n \geqslant 3/2$ and $F(2) = 360/49 = 7.3469\ldots$. The latter follows from $\zeta(4) = \pi^4/90$, $\zeta(6) = \pi^6/945$, and $\zeta(8) = \pi^8/9450$. Note that $\lim_{n\to\infty} F(n) = \pi^2 = 9.8696\ldots$.

To obtain the desired inequality for F note that for $p \geqslant 2$, $n \geqslant 3/2$ we have

$$\left(1 - \frac{1}{p^{2n}}\right)\left(1 + \frac{1}{p^{2n+1}}\right) < 1 - \frac{1}{p^{2n}} + \frac{1}{p^{2n+1}} \leqslant 1 - \frac{1}{p^{2n+1}} \leqslant 1 - \frac{1}{p^{4n-2}}.$$

Using the expression for ζ involving products and primes, we have

$$\begin{aligned}\zeta(4n-2)/\zeta(2n) &= \Pi(1 - 1/p^{2n})/(1 - 1/p^{4n-2})\\ &< \Pi 1/(1 + 1/p^{2n+1}) = \Pi(1 - 1/p^{2n+1})/(1 - 1/p^{4n+2})\\ &= \zeta(4n+2)/\zeta(2n+1),\end{aligned}$$

where the products run over all primes. Multiplying by $\zeta(4n)$, we have

$$\frac{\zeta(4n)\zeta(4n-2)}{\zeta(2n)} < \frac{\zeta(4n+2)\zeta(4n)}{\zeta(2n+1)} \quad \text{for } n \geqslant 3/2.$$

Applying this for n and $n + 1/2$ yields

$$\zeta(4n)\zeta(4n-2)/\zeta(2n) < \zeta(4n+2)\zeta(4n)/\zeta(2n+1) < \zeta(4n+4)\zeta(4n+2)/\zeta(2n+2).$$

Finally, noting that $\zeta(2n) > \zeta(2n+2)$, we obtain the desired inequality $F(n) = \pi^2\zeta(4n)\zeta(4n-2)/(\zeta(2n))^4 < \pi^2\zeta(4n+4)\zeta(4n+2)/(\zeta(2n+2))^4 = F(n+1)$.

Editorial comment. The monotonicity of the function F defined by Shafer can be obtained more simply by logarithmic differentiation, since

$$\frac{F'(s)}{F(s)} = -8\frac{\zeta'(2s)}{\zeta(2s)} + 4\frac{\zeta'(4s)}{\zeta(4s)} + 4\frac{\zeta'(4s-2)}{\zeta(4s-2)} = \sum_{n=1}^{\infty}\Lambda(n)\left(\frac{8}{n^{2s}} - \frac{4}{n^{4s}} - \frac{4}{n^{4s-2}}\right) > 0$$

for any real s greater than 1. Here $\Lambda(n) = \log p$ if n is a power of the prime number p and $\Lambda(n) = 0$ otherwise.

Several solvers noted that the problem is essentially solved by the inequality

$$\frac{2(2n)!}{(2\pi)^{2n}} < |B_{2n}| < \frac{2(2n)!}{(2\pi)^{2n}(1-2^{1-2n})}$$

in M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, National Bureau of Standards (1964), p. 805 eq. 23.1.15. This yields the desired inequality if $\pi^2(1-2^{1-2n})^4 \geqslant 1$, which is true for $n \geqslant 2$.

Also solved by N. Franceschine, C. Georghiou (Greece), R.-F. Gloden (Italy), C. Hill, O. P. Lossers (The Netherlands), C. Vandermee (The Netherlands), and the proposer.

Inverse of a Combinatorial Matrix

E 3161 [1986, 565]. *Proposed by Ira Gessel, Brandeis University, Waltham, MA.*

Let $A = (a_{ij})_{i,j\geqslant 0}$ be the infinite lower triangular matrix defined by $a_{ij} = \binom{i}{j}\binom{j}{i-j}$. Let $A^{-1} = (b_{ij})_{i,j\geqslant 0}$. Show that

$$\sum_{i,j=0}^{\infty} b_{ij}\frac{x^i}{i!^2}y^j = f((1+y)x)/f(x),$$

where

$$f(x) = \sum_{n=0}^{\infty}\frac{x^n}{n!^2} = I_0(2\sqrt{x}). \quad (I_0 \text{ is one of the modified Bessel functions.})$$

Solution by David Callan, University of Bridgeport. Multiply both sides of the desired equality by $f(x)$. To prove that the coefficients of x^ny^j on both sides are the same, we must show $\sum_{i=0}^{n}\binom{n}{i}^2 b_{ij} = \binom{n}{j}$. We can write this in matrices as $DB = C$, where $B = A^{-1}$, $C = [\binom{n}{j}]_{n,j\geqslant 0}$, and $D = [\binom{n}{i}^2]_{n,i\geqslant 0}$. Multiplying on both sides by A, the required equality holds if and only if $D = CA$, i.e., $\binom{n}{i}^2 = \sum_{j=0}^{n}\binom{n}{j}a_{ji}$.

Evaluating the sum using the Vandermonde convolution $\Sigma_i\binom{r}{i}\binom{s}{i} = \binom{r+s}{s}$, we have

$$\sum_{j=i}^{n}\binom{n}{j}\binom{j}{i}\binom{i}{j-i} = \binom{n}{i}\sum_{j=i}^{n}\binom{n-i}{j-i}\binom{i}{j-i} = \binom{n}{i}^2,$$

as desired.

Editorial comment. This problem was published with an unfortunate misprint, so that the matrix element a_{ij} was given as $\binom{i}{j}\binom{i}{i-j}$ instead of $\binom{i}{j}\binom{j}{i-j}$. For this misprinted problem,

$$\sum_{i,j=0}^{\infty} b_{ij}\frac{x^i}{i!^2}y^j = f(yx)/f(x),$$

as shown by J. M. Cohen, J. B. M. Melissen (The Netherlands), W. A. Newcomb, and G. S. Sylvester.

The intended problem was also solved by The Chico Problem Solving Group and the proposer.

A Series of Distances

E 3179 [1986, 812]. *Proposed by N. J. Lord, Tonbridge School, England.*

Given a real number x, let $f_n(x)$ denote the distance from x to the nearest rational number with denominator n (not necessarily in its lowest form). For which values of x does the series $\Sigma_{n=1}^{\infty} f_n(x)$ converge?

Solution by L. E. Mattics, University of South Alabama, Mobile, AL. If x is an integer the series converges to zero; we shall show that it diverges otherwise.

If x is irrational, then the numbers $nx - \lfloor nx \rfloor$ are uniformly distributed in the interval $(0, 1)$. (Cf. Hardy and Wright, *Theory of Numbers*, Th. 445.) Consequently, if A is the set of positive integers n such that $1/4 < nx - \lfloor nx \rfloor < 1/2$ and $A(m)$ is the number of integers n in A with $n < m$, then $\lim_{m\to\infty} A(m)/m = 1/4$. Hence the series $\Sigma_{n\in A} 1/(4n)$ diverges. But if $n \in A$, then $f_n(x) = x - \lfloor nx \rfloor/n \geqslant 1/(4n)$. Thus $\Sigma f_n(x)$ diverges.

Finally, suppose $x = r/s$, where r and s are relatively prime integers with $s > 1$, and let p be a prime not dividing s. Then $f_p(x) \geqslant 1/(ps)$. Since $\Sigma_{p\text{ prime}} 1/p$ diverges, $\Sigma f_n(x)$ also diverges.

Also solved by K. F. Andersen (Canada), E. H. Grossman, N. Felsinger, M. Gutkowski, O. P. Lossers (The Netherlands), W. A. Newcomb, and the proposer.

A Triangular Inequality

E 3180 [1986, 812]. *Proposed by Murray S. Klamkin, University of Alberta, Canada.*

If A, B, C are angles of a triangle, prove that

$$\cos\frac{A}{2} + \cos\frac{B}{2} + \cos\frac{C}{2} \geqslant 1 + \sin\frac{A}{2} + \sin\frac{B}{2} + \sin\frac{C}{2}.$$

Solution by O. P. Lossers, Eindhoven University of Technology, Eindhoven, The Netherlands. We consider the function V defined by

$$V(A, B, C) = \cos\frac{A}{2} + \cos\frac{B}{2} + \cos\frac{C}{2} - \left(\sin\frac{A}{2} + \sin\frac{B}{2} + \sin\frac{C}{2}\right).$$

on the compact set S defined by $A \geqslant 0$, $B \geqslant 0$, $C \geqslant 0$, and $A + B + C = \pi$. V attains an absolute maximum and minimum; we will show that the minimum is 1 and the maximum is $3(\sqrt{3} - 1)/2$.

Suppose $C = 0$, so that $A + B = \pi$. The identity $\cos(\pi/2 - x) = \sin x$ implies $V(A, B, 0) = 1$. Similarly $V(0, B, C) = V(A, 0, C) = 1$, so V is identically 1 on the boundary of S, and we need only consider interior points.

For fixed $C \neq 0$, we have

$$\begin{aligned} V(A, B, C) &= \left(\cos\frac{A}{2} + \cos\frac{B}{2}\right) - \left(\sin\frac{A}{2} + \sin\frac{B}{2}\right) + \left(\cos\frac{C}{2} - \sin\frac{C}{2}\right) \\ &= 2\cos\frac{A - B}{4} \cdot \left(\cos\frac{A + B}{4} - \sin\frac{A + B}{4}\right) + \cos\frac{C}{2} - \sin\frac{C}{2} \\ &= 2\cos\frac{A - B}{4} \cdot \left(\cos\frac{\pi - C}{4} - \sin\frac{\pi - C}{4}\right) + \cos\frac{C}{2} - \sin\frac{C}{2}. \end{aligned}$$

Since $|x| < \pi/4$ implies $\cos x > \sin x$, the last expression is minimized only when $A = 0$ or $B = 0$, and it is maximized when $A = B$. This shows that V is in fact minimized on the boundary $A \cdot B \cdot C = 0$. As for the maximum, by symmetry V is also maximized for a fixed $A \neq 0$ when $B = C$. Hence V attains its maximum only at $A = B = C = \pi/3$, where $V = 3/2(\sqrt{3} - 1)$. Thus we have $1 < V(A, B, C) \leqslant 3/2(\sqrt{3} - 1)$, with lower equality only at a degenerate triangle and upper equality only at an equilateral triangle.

Editorial Comment. Several readers used Lagrange multipliers to find the extrema of V. Lagrange's method seems particularly suitable for studying the extrema of such functions of the angles of a triangle. For example, it shows that the cosine sum and the sine sum separately satisfy the inequalities

$$2 < \cos\frac{A}{2} + \cos\frac{B}{2} + \cos\frac{C}{2} \leqslant \frac{3}{2}\sqrt{3},$$

$$1 < \sin\frac{A}{2} + \sin\frac{B}{2} + \sin\frac{C}{2} \leqslant \frac{3}{2},$$

with lower equality only for a degenerate triangle and upper equality only for an equilateral triangle. H. Guggenheimer remarks that the difference between the

maximum and minimum values of V is surprisingly small when compared to the corresponding differences in the inequalities just quoted.

Solved also by the proposer and 32 others.

A Convergent Sequence

E 3197 [1987, 300]. *Proposed by L. A. Rubel, University of Illinois, Urbana, IL.*

Suppose $a_1, a_2, a_3, \ldots$ is a sequence of real numbers satisfying $0 < a_1 \leqslant 1$ and $0 \leqslant a_n \leqslant 1$ for $n = 2, 3, \ldots$. Put $S_n = a_1 + a_2 + \cdots + a_n$ and $T_n = S_1 + S_2 + \cdots + S_n$. Must $\Sigma_{n=1}^{\infty} a_n/T_n$ converge?

Solution by L. E. Mattics, University of South Alabama. We show more generally that $\Sigma a_n/T_n^{\alpha}$ converges as long as $\alpha > 1/2$. If Σa_i converges, then the problem is trivial, since $1/T_n \leqslant 1/a_1$ for all n, so henceforth we assume Σa_i diverges.

Since $0 \leqslant a_n \leqslant 1$ and Σa_i diverges, for each integer $m > 0$ there is a unique positive integer $n(m)$ such that $S_{n(m)} > m \geqslant S_{n(m)-1}$. The sequence $n(m)$ is strictly increasing, so

$$T_{n(m)} = S_1 + \cdots + S_{n(m)} \geqslant S_{n(1)} + \cdots + S_{n(m)} > m(m+1)/2.$$

The sequence T_k is also strictly increasing, so

$$\sum_{n=n(2)}^{\infty} \frac{a_n}{T_n^{\alpha}} = \sum_{m=2}^{\infty} \sum_{k=n(m)}^{n(m+1)-1} \frac{a_k}{T_k^{\alpha}} \leqslant \sum_{m=2}^{\infty} \frac{\left(S_{n(m+1)-1} - S_{n(m)-1}\right)}{T_{n(m)}^{\alpha}}.$$

Because the parenthesized quantity cannot exceed 2, this sum is bounded by $2\Sigma_{m=2}^{\infty}[(m^2+m)/2]^{-\alpha}$, which converges for $\alpha > 1/2$.

Editorial Comment. Grahame Bennett remarked that setting $a_n = 1$ for all n proves that the result above is best possible, by providing a divergent example when $\alpha \leqslant \frac{1}{2}$.

Also solved by G. Bennett, E. Hertz, W. B. Johnson, O. P. Lossers (The Netherlands), A. Riese and G. Fricke, and the proposer.

ADVANCED PROBLEMS

6579. *Proposed by D. K. Cohoon, Temple University, Philadelphia, PA.*

Suppose we define a curl operation on the infinitely differentiable functions from R^7 to C^7 by the rule

$$\operatorname{curl}(\vec{f}) = \sum_{i=1}^{7} \left\{ \left(\frac{\partial f_{i+3}}{\partial x_{i+1}} - \frac{\partial f_{i+1}}{\partial x_{i+3}} \right) + \left(\frac{\partial f_{i+6}}{\partial x_{i+2}} - \frac{\partial f_{i+2}}{\partial x_{i+6}} \right) + \left(\frac{\partial f_{i+5}}{\partial x_{i+4}} - \frac{\partial f_{i+4}}{\partial x_{i+5}} \right) \right\} \vec{e}_i$$

where $\vec{e}_i$ is the unit vector in the positive direction on the ith coordinate axis, $\vec{f} = \vec{f}(x) = f_1\vec{e}_1 + f_2\vec{e}_2 + f_3\vec{e}_3 + f_4\vec{e}_4 + f_5\vec{e}_5 + f_6\vec{e}_6 + f_7\vec{e}_7$ is a C^∞ function from R^7 to C^7, and, for all positive integers j in $\{1,2,3,4,5,6,7\}$, $x_{j+7} = x_j$ and $f_{j+7} = f_j$. Show that every infinitely differentiable function from R^7 to C^7 is a curl plus a gradient.

6580. *Proposed by P. A. Batnik, University of Illinois at Urbana-Champaign.*

Prove that if n is an odd positive integer, there exist integers x_1, x_2, x_3, x_4 such that

$$n = x_1^2 + x_2^2 + x_3^2 + x_4^2, \qquad x_1 > \lfloor\sqrt{n}\rfloor - 1.$$

6581. *Proposed by Donald E. Knuth and Ilan Vardi, Stanford University.*

Stirling's approximation leads to the following expansion

$$\binom{2n}{n} = \frac{4^n}{\sqrt{\pi n}}\left\{1 - \frac{1}{8n} + \frac{1}{128n^2} + \frac{5}{1024n^3} - \frac{21}{32768n^4} + O\left(\frac{1}{n^5}\right)\right\}.$$

Prove that if this expansion is continued indefinitely, the coefficient of n^{-k} will be $2^{1-4k}m_k$, where m_k is an integer. What is the exact power of 2 that divides the denominator of this coefficient?

6582. *Proposed by N. J. Fine, Deerfield Beach, FL.*

Prove that

$$\int_0^{e^{-\pi}} \prod_{n=1}^{\infty} \frac{(1-q^{2n})^{20}}{(1-q^n)^{16}}\,dq = \frac{1}{16}.$$

SOLUTIONS OF ADVANCED PROBLEMS

Weak Contractions

6532 [1986, 822]. *Proposed by Victor Pambuccian, Bucharest, Romania.*

Prove or disprove the following "converse" of Edelstein's theorem (see this MONTHLY, Problem S 8 [1979, 222; 1980, 487]):

Let X be a set of power less than or equal to $2^{\aleph_0}$, and $f\colon X \to X$ a map such that each f^n $(n = 1,2,\dots)$ has a unique fixed point. Here f^n means $f\circ f\circ\cdots\circ f$ (n times). Then there exists a metric d on X such that (X, d) is a compact metric space and

$$d(f(x), f(y)) < d(x, y) \quad \forall x, y \in X, \qquad x \neq y.$$

Note that a corresponding "converse" holds for Banach's fixed point theorem ("contraction principle") (see C. Bessaga, *Colloq. Math.*, 7 (1959) 41–43).

Solution by the proposer. The "converse" is false. Let $X = N$, the set of positive integers, and define an integer valued function f by

$$f(m) = \begin{cases} 1 & m = 1 \\ m-1 & m \geqslant 2. \end{cases}$$

Each iterate f^n, $n = 1, 2, \ldots$, has $x = 1$ as its unique fixed point. Any metric d on X that satisfies

$$d(f(x), f(y)) < d(x, y), \quad \text{for all } x \neq y,$$

must clearly satisfy

$$d(1, k) < d(2, k+1) < d(3, k+2) < \cdots, \quad \text{all } k \geqslant 2,$$

and also

$$d(1,2) < d(1,3) < d(1,4) < \cdots.$$

Therefore,

$$d(1, k) < d(1+i, k+i), \quad \text{all } k \geqslant 2, i \geqslant 1$$

and

$$d(1,2) < d(1, k), \quad \text{all } k \geqslant 3.$$

This implies that

$$d(1,2) < d(i, j)$$

for all $i \neq j$ such that $\{i, j\} \neq \{1,2\}$. Let $r = d(1,2)$. Then

$$B_r(x) = \{y \in X: d(x, y) < r\} = \{x\} \quad \text{for all } x \in X.$$

In other words, the "open ball" $B_r(x)$ must consist solely of its center point x for any such metric d. Hence (X, d) is a discrete topological space and so is not compact.

Editorial Comment. Edelstein's theorem asserts that if (X, d) is a compact metric space and $f\colon X \to X$ is a mapping such that

$$d(f(x), f(y)) < d(x, y), \quad \text{all } x, y \in X, \qquad x \neq y,$$

then f has a unique fixed point. (Cf. M. Edelstein, *J. London Math Soc.*, 37 (1962) 74–79 or Victor Bryant, *Metric Spaces*, Cambridge, 1985, p. 69.) Since a compact metric space is a continuous image of the Cantor set, a converse would presumably involve a set of cardinality less than or equal to that of the continuum. Now C. Bessaga (*Colloq. Math.* 7 (1959), 41–43) has proved a converse of the Banach fixed point theorem for complete metric spaces whose formulation bears the same relation to Banach's theorem that the assertion of the problem bears to Edelstein's. From this perspective the counterexample seems a bit surprising. Thomas Jager points out that we can obtain a "converse" for X *finite*. In this case there is an n such that

$$X \supset f(X) \supset \cdots \supset f^n(X) = \{p\}$$

where p is the fixed point. For x in X and $x \neq p$, let $k(x)$ be that integer for which

$$x \in f^{k(x)}(X) - f^{k(x)+1}(X),$$

and set $k(p) = n$. Then the metric d defined by

$$d(x, y) = 2n - k(x) - k(y)$$

for $x \neq y$ and $d(x, x) = 0$ makes X into a compact metric space on which f is a contraction.

Thomas Jager also constructed a countable counterexample. Ludvik Janos provided an uncountable but quite simple counterexample with X the set of all real numbers and $f(x) = x/2$. The idea is to let x^* maximize $d(0, x)$, and then establish the contradiction

$$d(0, 2x^*) \leqslant d(0, x^*) = d(0, f(2x^*)) < d(0, 2x^*).$$

Borel Images

6534 [1987, 81]. *Proposed by F. S. Cater, Portland State University, Portland, Oregon.*

Let f and g be real-valued functions on the real line, $\mathbb{R}$, such that if B is any Borel subset of $\mathbb{R}$, then $f(B)$ and $g(B)$ are also Borel sets. For each number y, let $f^{-1}(y)$ be at most a finite set.

(a) Prove that there exist Borel sets A_1, A_2, B_1, B_2 such that

$$A_1 \cup A_2 = B_1 \cup B_2 = \mathbb{R}, \qquad A_1 \cap A_2 = B_1 \cap B_2 = \varnothing,$$
$$f(A_1) = B_1, \qquad g(B_2) = A_2.$$

(b)* Is the hypothesis on f^{-1} essential?

Solution of (a) by the proposer. The problem requires

$$f(R \setminus g(S)) = R \setminus S,$$

where $S = B_2$. Our strategy is to first create a sequence of sets S_n for which

$$f(R \setminus g(S_n)) = R \setminus S_{n+1}.$$

To do this, set $S_0 = R \setminus f(R)$ and

$$S_{n+1} = R \setminus f(R \setminus g(S_n)), \qquad n \geqslant 0.$$

Then take $B_2 = S$ where

$$S = \bigcup_{n=0}^{\infty} S_n.$$

Now we provide the details. Observe that S_0 is a Borel set, and an easy induction on n shows that S_n is a Borel set for all n. Thus S is a Borel set and so is $g(S)$.

Now

$$(R \setminus S_1) \cap S_0 = f(R \setminus g(S_0)) \cap (R \setminus f(R)) = \varnothing$$

so $S_0 \subset S_1$. It is clear that

$$S_1 = R \setminus f(R \setminus g(S_0)) \subset R \setminus f(R \setminus g(S_1)) = S_2$$

and by induction we see that

$$S_n = R \setminus f(R \setminus g(S_{n-1})) \subset R \setminus f(R \setminus g(S_n)) = S_{n+1}$$

for all $n \geqslant 1$. Thus $S_0 \subset S_1 \subset \cdots \subset S_n \subset \cdots$. But $S = \bigcup_n S_n$. So

$$f(R \setminus g(S)) = f(R \setminus \textstyle\bigcup_n g(S_n)) = f[\bigcap_n (R \setminus g(S_n))]. \tag{1}$$

But for any y, $f^{-1}(y)$ is at most a finite set. It follows from this and $g(S_n) \subset g(S_{n+1})$

that

$$f[\cap_n(R \setminus g(S_n))] = \cap_n f(R \setminus g(S_n)). \tag{2}$$

From (1) and (2) we obtain

$$f(R \setminus g(S)) = \cap_n f(R \setminus g(S_n)). \tag{3}$$

But $R \setminus S_{n+1} = f(R \setminus g(S_n))$ and it follows that

$$\cap_n f(R \setminus g(S_n)) = \cap_n(R \setminus S_{n+1}) = R \setminus \cup_n S_{n+1} = R \setminus S,$$

and

$$\cap_n f(R \setminus g(S_n)) = R \setminus S. \tag{4}$$

From (3) and (4) we obtain

$$f(R \setminus g(S)) = R \setminus S. \tag{5}$$

Put $A_1 = R \setminus g(S)$, $A_2 = g(S)$, $B_1 = R \setminus S$, $B_2 = S$. The conclusion, then, follows from (5).

No solutions to (b) were received. The editor thanks Michael J. Dixon for some helpful comments.

Absolute Integrability

6535 [1987, 81]. *Proposed by Brockway McMillan, Sedgwick, Maine.*

Let $D = [0, \infty) \times [0, 1)$ and $f = f(x, y)$ be a real-valued function absolutely integrable on bounded measurable subsets of D. Assume that

$$\lim_{k \to \infty} \iint_{A_k \times B_k} f(x, y)\, dx\, dy = 0$$

for every pair of ascending sequences $A_1 \subseteq A_2 \subseteq \cdots$, $B_1 \subseteq B_2 \subseteq \cdots$ of bounded measurable sets such that

$$\bigcup_1^\infty A_k = [0, \infty) \quad \text{and} \quad \bigcup_1^\infty B_k = [0, 1).$$

Must f be absolutely integrable over D?

Solution by M. J. Pelling, University College, London, England. No. For $n = 1, 2, \ldots$ define

$$f = f(x, y) = \begin{cases} 1/n & \text{if } n - 1 \leqslant x < n \text{ and } \dfrac{2k}{2^n} \leqslant y < \dfrac{2k+1}{2^n}, \\ & 0 \leqslant k \leqslant 2^{n-1} - 1, \\ -1/n & \text{if } n - 1 \leqslant x < n \text{ and } \dfrac{2k+1}{2^n} \leqslant y < \dfrac{2k+2}{2^n}, \\ & 0 \leqslant k \leqslant 2^{n-1} - 1. \end{cases}$$

Then

$$\iint_D |f|\, dx\, dy = \infty,$$

yet

$$\iint_{A_k \times B_k} f\,dx\,dy \to 0 \quad \text{as } k \to \infty.$$

The first statement is immediate, since the double integral of the modulus over D is

$$\sum_{n=1}^{\infty} \iint_{[n-1,\,n]\times[0,1]} |f|\,dx\,dy = \sum_{n=1}^{\infty} \frac{1}{n}.$$

For the second, we write each y in $[0,\ 1)$ as a binary "decimal" $y = 0.c_1c_2c_3\cdots$ (use the terminating expansion in the ambiguous cases) and set

$$\varepsilon_m(y) = \begin{cases} 1 & c_m = 0 \\ -1 & c_m = 1. \end{cases}$$

Then $f(x, y) = \varepsilon_m(y)/m$ for $m - 1 \leqslant x < m$ and

$$g_n(y) = \int_0^n f(x, y)\,dx = \sum_{m=1}^{n} \frac{\varepsilon_m(y)}{m}.$$

By orthogonality

$$\int_0^1 (g_{n+k}(y) - g_n(y))^2\,dy = \frac{1}{(n+1)^2} + \frac{1}{(n+2)^2} + \cdots + \frac{1}{(n+k)^2} \leqslant \frac{1}{n},$$

so by the Schwarz inequality

$$\int_0^1 |g_{n+k}(y) - g_n(y)|\,dy \leqslant \frac{1}{\sqrt{n}}.$$

Thus $\{g_n\}$ is a Cauchy sequence in the L_1 norm, and by completeness we obtain the existence of a real-valued L_1 function g on $[0,1]$ such that

$$\int_0^1 |g_n(y) - g(y)|\,dy \to 0 \quad \text{as } n \to \infty.$$

We are now ready to study the restricted double integrals. For any bounded (measurable) set A_k,

$$\iint_{A_k \times [0,1]} f\,dx\,dy = \int_{A_k} \left[\int_0^1 f\,dy \right] dx = 0,$$

since the inner integral is always zero. Hence

$$\iint_{A_k \times B_k} f\,dx\,dy = -\iint_{A_k \times B_k^c} f\,dx\,dy,$$

where B_k^c is the complement in $[0,1]$ of B_k. Therefore it suffices to show that for any $\varepsilon > 0$ there is a $\delta > 0$ such that for all bounded $A \subseteq [0, \infty)$ and all $S \subseteq [0,1]$ with $m(S) < \delta$ we have

$$\left| \iint_{A \times S} f\,dx\,dy \right| < \varepsilon.$$

Assume that A and S are sets as above. Let

$$I_n = m(A \cap [n-1, n]).$$

Then

$$\left|\iint_{A\times S} f\,dx\,dy\right| = \left|\sum_{m=1}^{\infty} I_m \int_S \frac{\varepsilon_m(y)}{m}\,dy\right|$$
$$\leqslant \sum_{m=1}^{\infty}\left|\int_S \frac{\varepsilon_m(y)}{m}\,dy\right|$$
$$\leqslant \left\{\sum_{m=1}^{\infty} m^{-2}\right\}^{1/2}\left\{\sum_{m=1}^{\infty}\left(\left(\int_S \varepsilon_m(y)\,dy\right)\right)^2\right\}^{1/2}.$$

If we let χ_S be the characteristic function of the set S, then the square of the second factor above is

$$\sum_{m=1}^{\infty}\left\{\int \varepsilon_m(y)\chi_S(y)\,dy\right\}^2$$

and, by Bessel's inequality, this is at most

$$\int \chi_S(y)^2\,dy = m(S).$$

Thus, whenever $m(S)$ is small, so is

$$\iint_{A\times S} f\,dx\,dy.$$

This completes the proof.

Editorial Comment. Pelling's solution is, of course, an application of Rademacher functions. For other nice things that can be done with them, see M. Kac, *Statistical Independence in Probability, Analysis, and Number Theory*, Math. Assn. of Amer. (Carus Monograph Number Twelve), 1959. The proposer's original statement of the problem involved n-fold integrals, and he expected the problem to be answered with a counterexample. As for its origin, he informs us that it "is abstracted from work on statistical mechanics, where one often meets infinite series of which the nth term is [related to the sort of n-fold] multiple integral appearing in the statement of the problem."

Automorphisms of the Symmetric Group on Six Symbols

6538 [1987, 195]. *Proposed by Joseph Rotman, University of Illinois, Champaign-Urbana.*

Let $\phi \in \mathrm{Aut}(S_6)$ have the following values on transpositions:

$$\phi(12) = (12)(36)(45)$$
$$\phi(13) = (16)(24)(35)$$
$$\phi(14) = (13)(25)(46)$$
$$\phi(15) = (15)(26)(34)$$

Find $\phi(16)$ explicitly.

Combined solution. Since ϕ is an automorphism, it must preserve conjugacy classes, and so $\phi(16)$ is a product of three disjoint transpositions. If $\phi(16)$ shares a transposition with $\phi(1i)$ for some i with $2 \leqslant i \leqslant 5$, then it is easy to check that $\phi(16)$ commutes with $\phi(1i)$. This is a contradiction since (16) and $(1i)$ do not commute. Hence $\phi(16) = (14)(23)(56)$.

Editorial remark. In G. J. Janusz and J. J. Rotman, Outer Automorphisms of S_6, this MONTHLY, 89 (1982) 407–410, the authors constructed an outer automorphism ϕ and computed the values $\phi(1i)$ for $2 \leqslant i \leqslant 5$ (as in this problem) with no difficulty. However, in light of the solution to this problem, their computation of $\phi(16)$ is overcomplicated.

Also solved by F. Rudolf Beyl, S. H. Cullinane, John Dalbec (student) and S. F. Barger (jointly), Jesús Ferrer (Spain), Steven M. Gagola, Jr., the late I. N. Herstein, A. A. Jagers (The Netherlands), Keith A. Kearnes, O. P. Lossers (The Netherlands), David E. Manes, Ondřej Matouš (Czechoslovakia), John Henry Steelman, Joseph Verret, William P. Wardlaw, and the proposer.

REVIEWS

EDITED BY JOSEPH KONHAUSER
Department of Mathematics, Macalester College, St. Paul, MN 55105

A History of Algebra. By B. L. van der Waerden, Springer-Verlag, 1985. iii +271 pp.

DAVID J. WINTER
Department of Mathematics, University of Michigan, Ann Arbor, MI 48107

In *A History of Algebra*, Bartel Leenert van der Waerden covers the period from Muhammad ben Musa al-Khwarizmi to Emmy Noether, that is, the period from about the beginning of the 9th century to the first third or so of the 20th century.

A History of Algebra is written by a mathematician for mathematicians. In it, van der Waerden emphasizes the development of the mathematics and the interrelationships among ideas coming from different mathematicians. He discusses what was or may have been the motivation for research in a particular direction, as well as other factors and influences. He shows what mathematics was discovered and what the work really was about by describing the mathematics and showing how it was done —sometimes giving the crucial ideas, sometimes sketching a proof or outlining a manuscript, sometimes including proofs themselves.

The style is at the same time informal and precise. Although substantial mathematical prerequisites constrain the potential readership, *A History of Algebra* should be of interest to a fairly broad spectrum of readers from the mathematical community.

For the working algebraist or the student of algebra, *A History of Algebra* should be particularly useful. Van der Waerden not only places very important parts of algebra in historical perspective, but also gives an excellent expository discussion of some of the mathematics itself. He covers many important topics at length and, throughout the book, gives sources for material covered elsewhere. Furthermore, he treats much material which algebra students normally do not see in their graduate training, material which is important in understanding how and why the mathematics was created but is not needed in the modern logical development of it. At the same time, such material can provide deep insight into the modern theory and thus both render the theory easier to understand and clarify its significance and potential future.

One notable asset of *A History of Algebra* is the fact that the subject matter is not at all confined to algebra. The author is concerned not only with the algebra, but also with the mathematicians who discovered it and with their other related work, with the surrounding circumstances and with the relations among their various discoveries. It was therefore practical and necessary to include a wide variety of interesting topics from geometry, topology, analysis, physics, astronomy and other subjects, which are well integrated into the discussion.

Van der Waerden's choice of material, and his balance between the emphasis on history and emphasis on the mathematics, is reasonable. At the same time, algebra is a field about which much more can be said than can be put in any one book. Accordingly, there is a need for other books covering other important topics and other aspects of its history.

A History of Algebra comprises three parts, divided according to subject. Part I: Algebraic Equations occupies half of the book. Part II: Group Theory and Part III: Algebras share the second half. Of course, this division is ignored when, for instance, groups are used to study equations. So, considerable attention is given in Part I to the group theory of Cauchy, Galois, and Jordan, that is, to group theory as it was before Kronecker, Cayley, and others developed the abstract notion *group*. For Cauchy, Galois, and Jordan, a finite group was a finite set of *substitutions* (permutations) closed under multiplication, and group theory was a powerful tool for studying equations and, for Jordan, transcendental functions.

In an attempt to describe what is done in *A History of Algebra*, we look briefly at each of the three parts and sample some of the material.

In Part I, van der Waerden traces the development of the theory of algebraic equations from Muhammad ben Musa al-Khwarizmi, in the first half of the 9th century, to Evariste Galois (1811–1832) and Camille Jordan (1838–1922).

Muhammad ben Musa al-Khwarizmi worked with equations such as "A square, which is equal to forty things minus four squares," which we represent in present day notation by $x^2 = 40x - 4x^2$, and with algebraic operations such as *al-jabr* or *al-muqabala* (adding or subtracting equals to or from equals). According to the biographer Haji Khalfa, al-Khwarizmi was the first Islamic author to write on solving equations by performing such algebraic operations.

Al-Khwarizmi was a member of the *House of Wisdom* in Bagdad, under the Caliph al-Mamum, who reigned in the period A.D. 813 to A.D. 833. His treatise *Algebra*, dedicated to the Caliph, energetically ignored Euclid's *Elements*, even though it had been translated into Arabic by his colleague al-Hajjaj ibn Yusuf ibn Mater. Solomon Gandz, the editor of the *Mishnat ha-Middot*, an ancient Hebrew treatise on mensuration, perhaps written by Rabbi Nehemiah about A.D. 150 and being, perhaps, the primary source used by al-Khwarizmi, writes:

> On the contrary, in the preface to his *Algebra* al-Khwarizmi distinctly emphasizes his purpose of writing a popular treatise that, in contradiction to Greek theoretical mathematics, will serve the practical ends and needs of the people in their affairs of inheritance and legacies, in their lawsuits, in trade and commerce, in the surveying of lands and in the digging of canals. Hence, al-Khwarizmi appears to us not as a pupil of the Greeks but, quite to the contrary, as the antagonist of al-Hajjaj and the Greek school, as the representative of the native popular sciences.

Al-Khwarizmi explains how all linear and quadratic equations can be reduced to one of the six forms

$$ax^2 = bx, \qquad ax^2 = b, \qquad ax = b, \qquad ax^2 + bx = c,$$
$$ax^2 + c = bx, \qquad ax^2 = bx + c.$$

He then gives rules for solving each kind, and illustrates them by worked examples.

Van der Waerden then discusses the writings of al-Khwarizmi on computing areas, on the Jewish calendar, on legacies, on geography, on Hindu numerals and on astronomy. Finally, he considers the question of what sources may have been used by al-Khwarizmi.

Van der Waerden discusses two other early Muslim mathematicians. One of these, Omar Khayyam, best known for his 600 or so poems in the *Rubaiyat*, gave

systematic methods for solving equations. For linear and quadratic equations, he followed the geometrical methods explained in Euclid's *Elements*. He then showed how cubic equations can be solved by means of intersections of conics. Van der Waerden develops some of the surrounding circumstances and discusses in detail the methods for solving cubic equations.

The author then goes on to discuss important work of some early Italian mathematicians. He starts with the many and varied contributions from 1200 to 1225 of Leonardo da Pisa (also known as filio Bonnacci or Fibonacci), who was the son of Bonnacci, a secretary of the republic of Pisa. Later on, he gives the interesting history of the discovery of algebraic methods for solving cubic biquadratic equations discovered by Scipione del Ferro, Niccolo Tartaglia, Gerolamo Cardano, Lodovico Ferrari, and Rafael Bombelli.

With improvements in notation by Francois Vieté (1540–1603) over the cumbersome notation used by early mathematicians, a decimal system created in 1585 by Simon Stevin, the restoration and improvement of the coordinate methods and theory of conic sections of Apollonius by Pierre de Fermat (1601–1665) and the coordinate systems of René Descartes (1596–1650), the stage was set for the more sophisticated algebra and theory of equations that was soon to follow.

Van der Waerden traces these developments giving interesting details, such as Descartes' solution, using his coordinate systems, of the problem of Apollonius of determining the "locus of three or four lines."

After devoting a chapter to Edward Waring, Alexandre-Théophile Vandermonde, Joseph Louis Lagrange, Gianfransesco Malfatti, Paolo Ruffini, Augustin Louis Cauchy, and Niels Henrik Abel, he discusses the work of Carl Friedrich Gauss on the cyclotomic equation $x^m - 1 = 0$ and outlines three of his proofs of the "Fundamental Theorem of Algebra."

The author concludes Part I with two nice chapters on Evariste Galois and Camille Jordan.

Abel had already shown that the equation of the fifth degree cannot be solved by radicals. Galois, during the year before his tragic death in a duel on May 30, 1832 at the age of 20, worked out his famous correspondence between groups and fields and used it to establish criteria for the solvability of an equation by radicals in terms of groups. The night before the duel, he explained the fundamental ideas in a letter to Auguste Chevalier, together with a comment and a request:

> *Je me suis souvent hasardé dans ma vie à avancer des propositions dont je n'étais pas sûr; mais tous ce que j'ai écrit là est depuis bientôt un an dans ma téte, et il est trop de mon intéret de ne pas me tromper pour qu'on me soupconne d'énoncer des théorèmes dont je n'aurais pas la démonstration complète.*
>
> *Tu prieras publiquement Jacobi et Gauss de donner leur avis, non sur la vérité, mais sur l'importance des théorèmes.*

Galois also introduced the important concept of primitive groups of substitutions and showed that solvable primitive groups are p groups. His discussion for those of order p^2 ends with the words "C'est ce que je vais rechercher."

It was Jordan (1838–1922) who later carried out this research, in part of his 667 page *Traité des substitutions et des équations algébriques*. Among his many important contributions, Jordan completed and refined Galois' theory of equations and

applied it to the study of transcendental functions. He also extended Galois' theory of primitive substitution groups and classified the solvable ones. Using this classification, he went on to construct all solvable transitive groups of substitutions on d letters for any d. This he did by finding maximal solvable irreducible subgroups of certain linear groups, using recursive methods.

In Part II, Van der Waerden concentrates on the theory of groups, noting that

> After the appearance of Jordan's *Traité* in 1870, a fundamental change of character of the theory of groups took place. Before 1870, only two kinds of groups were considered, namely groups of substitutions (or permutations) and groups of geometrical transformations. After 1870, the abstract notion *group* was developed in several steps notably by Kronecker (1870), Cayley (1878), von Dyck (1882) and Weber (1882).

He divides his discussion into four parts: Groups of Substitutions, Groups of Transformations, Abstract Groups, The Structure of Finite Groups.

After briefly discussing early theorems on the symmetric group, two papers by E. Mathieu on multiply transitive groups of substitutions and the structure theorems of the Norwegian mathematician M. L. Sylow, van der Waerden discusses groups of transformations. Whereas the work of Euler, Rodrigues and Jordan on closed groups of rigid motions of 3-space was discussed in Part I, the author now takes up the work of Arthur Cayley and Felix Klein on non-Euclidean geometry, Klein's classification of groups of fractional linear transformations and Sophus Lie's work on continuous groups.

Cayley develops a theory of distances having special cases leading to the *Euclidean plane* and to the *elliptic plane*. Klein then builds a model of the *hyperbolic geometry* of Lobatchewsky and Bolyai. In his second paper on this work, in 1873, Klein introduces the group generated by Euclidean displacements, similarity transformations and reflections. His fundamental idea is that each method of geometry is characterized by a transformation group.

Much of the remainder of the theory of finite groups of the nineteenth century was due to Otto Hölder. Whereas Jordan knew that the indices in a composition series for a finite group are unique up to their order, Hölder defined the concept *factor group* and showed that the composition factors of a composition series are unique up to their order.

One of the problems considered by Hölder was concerned with determining the structure of a group G from that of a normal subgroup H and the quotient group G/H. Later on, methods for solving this problem were developed by Otto Schreier.

In the last chapter of Part II, van der Waerden discusses the work of Sophus Lie, Friedrich Engel, Wilhelm Killing, and Elie Cartan culminating in the rich structure theory and classification theory of Lie groups and Lie algebras.

In Part III, van der Waerden discusses the discovery of algebras, beginning with the complex numbers. Complex numbers were, perhaps, first used by Cardano, who experienced "mental torture" in using them to express a number such as 40 as a product of numbers such as $5 + \sqrt{-15}$ and $5 - \sqrt{-15}$. Twenty-seven years after the publication of Cardano's *Ars Magna* (1545), they were used by Rafael Bombelli, who gave rules for calculating with them.

Van der Waerden then discusses Hamilton's discovery of the quaternions (and his long and fruitless search earlier for triplets) and the discovery by John Graves in 1843 of the octonions (and their rediscovery by Arthur Cayley in 1845). In 1853, Hamilton introduced the biquaternions

$$\begin{bmatrix} a+bi & -c+di \\ c+di & a-bi \end{bmatrix}.$$

Twenty years later, William Kingdom Clifford introduced two more kinds of biquaternions, one of which he (and, later, Eduard Study) used to study the rigid motions of Euclidean space.

In 1854, Cayley introduced the notion of group algebra. In 1862, Hermann Grassman introduced the Grassmann algebra. In 1878, Clifford introduced a variation of it, the *first Clifford algebra*, of 2^n dimensions, which is the quaternion algebra if n is 2. He also introduced an important subalgebra of it, the *second Clifford algebra*, which R. Lifschitz used six years later to represent rotations in n dimensions.

Van der Waerden next discusses the development of the structure theory of algebras, starting with the work of Benjamin Peirce, who discovered left, right and two-sided decompositions which are important for the later theory of Wedderburn.

Later on, he discusses Theodor Molien (1861–1941), who showed that every finite dimensional simple algebra over $\mathbb{C}$ is a complete matrix algebra. Cartan also proved this theorem and showed furthermore that every finite dimensional algebra over $\mathbb{C}$ is a direct sum of a simple or semisimple subalgebra and a nilpotent ideal.

Van der Waerden then briefly discusses the general theory of J. H. Maclagan Wedderburn, which generalizes the above results of Cartan on finite dimensional algebras over $\mathbb{C}$ to finite dimensional algebras over any field. He then briefly discusses work by Emil Artin and work by Emmy Noether.

He then discusses the structure theory of Nathan Jacobson for rings without finiteness assumptions.

Since the finite dimensional simple algebras are full matrix algebras over finite dimensional division algebras, to find them amounts to finding all finite dimensional division algebras D over fields F. Van der Waerden discusses how this is done for the following fields F:

the field $\mathbb{R}$
finite fields
p-adic fields
algebraic number fields.

If F is the real field, D is $\mathbb{R}$ or $\mathbb{C}$ or the quaternions, as proved by George Frobenius in 1878. If F is a finite field, then D is F, as proved by Wedderburn in 1909. For F a p-adic field, the structure of D was determined by Helmut Hasse in work completed in 1933. Using this and a local-global theorem for full matrix rings, Hasse, Brauer and Noether determine the structure of normal finite dimensional division algebras over algebraic number fields.

The remaining three chapters of the book are devoted to Group Characters, Representations of Finite Groups and Algebras and Representations of Lie Groups and Lie Algebras.

A First Course in Coding Theory. By Raymond Hill. Clarendon Press, Oxford, 1986, xii + 251 pp.

RICHARD HAMMING
Naval Postgraduate School, Monterey, CA 93940

It is often said (and the author of the present book says it once again) that coding theory arose from information theory. On the other hand, the person generally accepted as the founder of information theory is Claude Shannon, and Shannon points out in his seminal first paper on the subject that his initial results in coding theory were found by a method due to the reviewer, who thought he was working in coding theory! There were, in fact, two people working in the same place (Bell Telephone Laboratories), at about the same time, even sharing an office for awhile, who were developing two separate theories. (This of course raises the question of whether it is the situation that brings forth the creators or the creators that make the situation. One is reminded of Pasteur's remark that "luck favors the prepared mind," which admits both that there is an element of luck and that it is the individual who finally accomplishes the task.)

What was the motivation? Information theory arose from the desire to understand what the telephone system was doing in a fundamental sense. Coding theory, on the other hand, arose from the perceived need for reliable computing from unreliable parts, plus the need to signal through high levels of noise. Coding theory, therefore, arose in response to an important practical need, and the subject is still partially driven by practical considerations. This does not, however, explain the continued separation of the two subjects. That separation has persisted largely because the mathematics in the two fields seems to be somewhat different. Coding theory uses a great deal of finite algebra and combinatorics: Information theory tends to have much more classical analysis with such things as entropy functions and probability distributions. There are, of course, a few people who work in both fields, but the separation is quite noticeable.

While coding theory arose in response to practical needs, the typical treatment, as in this book, presents coding theory as another branch of mathematics. Usually, only slight attention is given to practical applications (other than *claims* of usefulness). Engineering realities are seldom discussed, and this book is no exception in this regard.

Considered as a branch of mathematics, just how hard is coding theory? Many authors of books on coding theory state that for their book one needs only simple high school algebra; the reader does *not* even need calculus. This is true in a formal sense, but it requires a mathematical sophistication that is not typically achieved before the calculus course. Hill writes, "The aim of this book is to provide an elementary treatment of the theory of error-correcting codes, assuming no more than high school mathematics and the ability to carry out matrix arithmetic... The first eight chapters comprise an introductory course which I have taught as part of second year undergraduate courses in discrete mathematics and in algebra. [This is less than half the book.] I have also used the whole as a master's course taken by students whose first degree is not necessarily in mathematics."

In the book, the author occasionally refers to the classic book by MacWilliams and Sloane (*The Theory of Error-Correcting Codes*, North-Holland, 1977) for some

of the harder proofs. Even so, his book is not easy reading for the person who is not mathematically sophisticated. It is well known that the Elementary Problems in the Monthly are often much harder than the Advanced Problems: combinatorial arguments are often very much harder than even involved analytical proofs. That the teacher, having mastered the field of finite algebra and the corresponding ways of thinking, finds the material easy does not mean that the student, meeting it for the first time, finds it easy. I can only wonder at the difficulty of covering the first one hundred pages of the book in one term. While one may cover hundreds of pages in the typical calculus text during the first term, such texts have voluminous sections of worked examples and carefully honed textual material. The job here cannot be easy! The author does give numerous examples to illustrate the material, and he does proceed carefully. He also gives background material such as the necessary parts of the theory of finite fields. Yet, with all this the subject material is not simple to grasp for the first time.

Given that the author views coding theory as a branch of mathematics, the book is excellent. Given that the field has many practical applications and is driven by them, there are many acknowledged omissions. (Huffman coding, for example, is only mentioned; the book is exclusively on block-coding.)

It is hard, however, to write a book covering coding theory that is easy to read and at the same time covers the more important codes; the field is too vast and difficult. From the mere note in Shannon's first paper, Golay understood a lot and wrote a classic one page paper that began the algebraic part of coding theory. It was a very difficult paper to read, and this tradition has persisted to this day. The field is difficult, but it is fair to say that the author of this book has made some further progress in simplifying the presentation of much of the material. For the calculus, a number of generations of teachers were needed to make the material simple enough for the average undergraduate to understand. In time, we will probably simplify the presentation of coding theory similarly.

As a branch of mathematics, coding theory has a promising future. There are many unsolved problems, and this feature is the mark of an important field of research. As a branch of practical engineering, it will receive many new stimulating problems in the future. The combination makes it a "hot field" for research; this book is one avenue of approach for the beginner.

TELEGRAPHIC REVIEWS

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T: Textbook	P: Professional Reading	1–4: Semesters
C: Computer Software	L: Undergraduate Library	**: Special Emphasis
S: Supplementary Reading	13: Grade Level	??: Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

General, S(15-16), P, L**.** *Mathematics and the Unexpected.* Ivar Ekeland. U of Chicago Pr, 1988, xiii + 146 pp, $19.95. [ISBN: 0-226-19989-4] A stimulating, selective account of a single mathematical idea—dynamical systems—traced from Kepler through Poincare to Arnold, Smale, and Thom. Using only words and diagrams, the author explains how deterministic mechanics can lead to random motion, and relates these two grand metaphors of science to ancient world views reflected by Homer. Translated by the author from an award-winning popular French monograph addressed to laymen. LAS

General, S(13-16), P, L**.** *The Mathematical Tourist: Snapshots of Modern Mathematics.* Ivars Peterson. WH Freeman, 1988, xv + 240 pp, $17.95. [ISBN: 0-7167-1953-3] A "Baedeker's Guide" to the sights and insights of modern mathematics adapted from reports written by the author for *Science News* during the past six years. Landmarks like primes, manifolds, fractals, knots, tiles, and cellular automata are linked by vivid, accessible exposition that conveys the ecstasy (without the agony) of many recent discoveries. A superb example of popular mathematical writing. LAS

Elementary, T(13: 1). *Algebra for College Students.* Margaret L. Lial, Charles D. Miller. Scott Foresman, 1988, 714 pp, $24.76. [ISBN: 0-673-18866-3] A substantial review of intermediate algebra precedes traditional college algebra topics; focuses on functions and problem solving; uses two colors even in equations. Available supplements include videotapes and audiotapes. JNC

Mathematics Appreciation, S(13-14), L*. *Discovering Mathematics: The Art of Investigation.* A. Gardiner. Oxford U Pr, 1987, xiv + 206 pp, $19.95 (P); $35. [ISBN: 0-19-853265-2; 0-19-853282-2] Unique book designed to give students experience in mathematical discovery. Four short and two extended investigations consist of a structured sequence of exercises interspersed with text. Topics from number theory, two-person games, geometry. Includes solutions for exercises and further problems. Appropriate for good high school students. KS

Finite Mathematics, T(13: 1). *Finite Mathematics and Its Applications.* Stanley J. Farlow, Gary M. Haggard. Random House, 1988, xi + 615 pp, $24. [ISBN: 0-394-35161-4] Covers linear equations, matrices, linear programming, elementary counting techniques, statistics, and graph theory. Applications are abundant (mathematics of finance, probability, etc.), and historical notes are interspersed throughout. Prerequisite is two years of high school algebra. LC

Education, P, L*. *Educating Scientists and Engineers, Grade School to Grad School.* Dir: John H. Gibbons. Office of Technology Assessment (US Government Printing Office, Washington, DC 20402), 1988, vii + 128 pp, $6 (P). A detailed analysis of patterns and policy issues pertaining to the flow of prospective science and engineering students through the educational pipeline. Major conclusion: demographic facts alone do not determine either the size or quality of the future science and engineering talent pool since there are proven ways to increase student entry into this pool, even as late as the sophomore year in college. Mathematics is generally aggregated with other sciences, so separate inferences about the mathematical pipeline are not easy to draw. LAS

Education, P*, L*. *The Mathematics Report Card: Are We Measuring Up?* John A. Dossey, *et al.* Educational Testing Service, 1988, 143 pp, (P). [ISBN: 0-88685-072-X] Summary and analysis of the National Assessment of Educational Progress (NAEP), a survey of mathematical knowledge of 9, 13, and 17 year olds conducted every four years. Major findings: Blacks and Hispanics have made appreciable gains, and average performance improved

slightly. However, most gains are in lower-order skills: only 6.4% of the 17 year olds could solve simple multi-step problems. Most students perceive mathematics as composed mainly of memorizing rules, and expect to have little use for mathematics in their lives. LAS

Education, P, L. *The Teacher of Mathematics: Issues for Today and Tomorrow.* Mathematical Sciences Education Board, (National Research Council, 2101 Constitution Avenue NW, Washington, DC 20418), 1987, v + 120 pp, (P). Papers presented at an October 1987 conference sponsored at UCLA by the Mathematical Sciences Education Board setting forth several related visions of the mathematics teacher of tomorrow. Includes major papers by Thomas Cooney on teaching practice, and by Albert Shanker on visions of new schools. LAS

Education, P*, L*. *A Conspiracy of Good Intentions: America's Textbook Fiasco.* Harriet Tyson-Bernstein. Council for Basic Education (725 15th St. NW, Washington, DC 20005), 1988, xii + 113 pp, (P). A scathing indictment of how public pressure to increase test scores and official regulation of the content of textbooks has led unintentionally to "dumbing down" school texts in which much is mentioned but little is taught. Concludes with a series of practical recommendations for how parents, policy makers, and educators can combat this tendency of bad texts to drive out good ones. LAS

Combinatorics, P. *Combinatorial Design Theory.* Ed: Charles J. Colbourn, Rudolf Mathon. Annals of Disc. Math., V. 149. North-Holland (US Distr: Elsevier Science), 1987, xii + 470 pp, Dfl. 170.00 (P). [ISBN: 0-444-70328-4] A collection of forty-one research papers ranging over the field: Latin squares, projective planes, Steiner systems, block designs, packings, coverings, graphs, etc., including attention to applications, computational tools, and construction problems. SS

Algebra, L. *Reviews in Ring Theory 1980-84.* AMS, 1986, xiii + 685 pp, $88 (P). [ISBN: 0-8218-0097-3] All reviews appearing in *Mathematical Reviews* from 1980 through 1984 with primary or secondary classification 16 (Associate Rings and Algebras). Subsections arranged by their MR numbers. Includes author index. LCL

Real Analysis, T(16-17), L. *Measure and Integral: Volume 1.* John L. Kelley, T.P. Srinivasan. Grad. Texts in Math., V. 116. Springer-Verlag, 1988, x + 150 pp, $39.80. [ISBN: 0-387-96633-1] Intended for use as a text for students with no previous knowledge of measure theory or Lebesgue integration, but augmented at the end of each chapter by informal discussions (in their totality comprising one-third of the book) of extensions of the ideas into more sophisticated mathematics. Intended as the first of a two-volume set, the second to consist of exercises, problems, and additional supplementary material. AWR

Differential Equations, T(17: 2). *Asymptotic Expansions for Ordinary Differential Equations.* Wolfgang Wasow. Dover, 1987, ix + 374 pp, $8.95 (P). [ISBN: 0-486-65456-7] An unabridged Dover republication of the edition published by Robert E. Krieger in 1976 (1965 Interscience edition, TR, January 1967; 1976 edition TR, August-September 1977). PS

Partial Differential Equations, T*(16-17). *Solutions of Partial Differential Equations.* Dean G. Duffy. TAB Books, 1986, x + 542 pp, $25.95. [ISBN: 0-8306-0412-X] A very nice book for a senior level course in partial differential equations. Starts with Sturm-Liouville problem, then covers Fourier series and integrals, Laplace transforms, convolution, and then applies these to first-order equations. Does lots of practical equations, and is method-oriented. MZ

Partial Differential Equations, P. *Computation of Singular Solutions in Elliptic Problems and Elasticity.* D. Leguillon, E. Sanchez-Palencia. Wiley, 1987, 200 pp, $42.95 (P). [ISBN: 0-471-91757-5] Presents general theory plus two new methods for computing the eigenvalues and eigenvectors of singularities. Concerns mostly homogeneous equations and boundary conditions. Many examples. Applications to engineering include boundary layers in composite plates and slightly perturbed corners. Discusses open problems. DFA

Operator Theory, P. *Proceedings of the Conference Ergodic Theory and Related Topics II.* Ed: Horst Michel. Teubner-Texte zur Math., B. 94. BG Teubner, 1987, 216 pp, 22,50M (P). [ISBN: 3-322-00434-1] Report of a conference held in 1986, this being a follow-up to a 1981 conference on the same theme. AWR

Functional Analysis, T*(18: 2), L. *Introduction to Functional Analysis, Second Edition.* Angus E. Taylor, David C. Lay. Robert E Krieger, 1986, xi + 467 pp, $42.50. [ISBN: 0-89874-951-4] A very readable text on the basics of functional analysis. It starts out by studying n-dimensional spaces, Zorn's lemma, and Hamel bases; it then treats normed linear spaces from the perspective of topological linear spaces, with Hilbert spaces and Banach spaces as special cases. In the second half it covers spectral analysis of linear operators in general and over Hilbert space. (1980 *Second Edition,* TR, December 1980.) MZ

Functional Analysis, T(16-17: 1, 2), L*. *Distribution Theory and Transform Analysis: An Introduction to Generalized Functions, with Applications.* A.H. Zemanian. Dover, 1987, xii + 371 pp, $8.95 (P). [ISBN: 0-486-65479-6] Republication of the 1965 edition in McGraw-Hill International Series in Pure and Applied Mathematics. Introduction to distribution theory and generalized Fourier and Laplace transformations. Applications to integro-differential equations and difference equations. Prerequisite is advanced calculus including familiarity with the interchange of limiting processes. Numerous examples, problems, and references (to 1965). JK

Algebraic Geometry, P. *Algebraic Geometry, Bowdoin 1985.* Ed: Spencer J. Bloch, *et al.* Proc. of Symp. in Pure Math., V. 46. AMS, 1985, $119

set [ISBN: 0-8218-1481-8]. *Part 1*, vi + 481 pp; *Part 2*, vi + 513 pp. Fifty-one articles reporting recent progress on the basic questions of algebraic geometry, many by leaders in the field. The three-week conference ranged from the algebraic (classification theorems, group actions, and algebraic cycles) to the geometric (vector bundles and Hodge theory). GG

Geometry, S(16-17), P, L. *Theory of Convex Bodies.* T. Bonnesen, W. Fenchel. BCS Associates, 1987, ix + 172 pp, $23 (P). [ISBN: 0-914351-02-8] This classic book is always cited as a source of basic results relating to convex bodies, but up to this time it has not been available in English. It would be an ideal addition to any library wanting an accessible, basic source on this topic. A classic! AWR

Geometry, T(18: 2), S, P. *Kleinian Groups.* Bernard Maskit. Grund. der math. Wissenschaften, B. 287. Springer-Verlag, 1988, xiii + 326 pp, $77.50. [ISBN: 0-387-17746-9] Intended for a one-year text at the graduate level, this is an essentially self-contained introduction to Kleinian groups as the theory has grown out of the initial work of Ahlfors and Bers. Topics include discontinuous groups, isometries, geometric and geometrically-finite groups, *B*-groups, and function groups. Exercises, bibliography. JS

Differential Topology, P.** *Collected Papers, Marston Morse, V. 1-6.* Marston Morse. World Scientific, 1987, $322 set. [ISBN: 9971-978-94-6] *V. 1*, xlix + 532 pp; *V. 2*, vi + 598 pp; *V. 3*, vi + 639 pp; *V. 4*, vi + 602 pp; *V. 5*, vi + 578 pp; *V. 6*, vi + 559 pp. Seven volume collected works of Marston Morse. Contains over 50 papers concerning Morse theory. Morse theory is a study of the relationship between the homology of a manifold and the critical points of its smooth functions. These volumes trace the development of Morse theory and its application to existence theorems for closed geodesics and to calculus of variations in general. Other topics covered include dynamics, minimal surface, topological methods in a single complex variable, integral representations, pseudoharmonic functions, and differential topology in papers written with collaborators such as G.A. Hedlund, C.B. Tompkins, M. Heins, W. Trunsue, J. Jenkins, W. Hubsch, and S.S. Cairns. AM

Optimization, P. *Design and Analysis of Algorithms for Stochastic Integer Programming.* L. Stougie. CWI Tract, V. 37. Math Centrum, 1987, 89 pp, Dfl. 14.10 (P). [ISBN: 90-6196-319-2] Treats hierarchical scheduling, vehicle, and location problems; stochastic integer programming by dynamic programming. The style is terse, the notation at times forbidding. AWR

Optimization, S(16-17), P. *Classical Principles and Optimization Problems.* B.S. Razumikhin. Math. & Its Applic. D Reidel (US Distr: Kluwer Academic), 1987, xiv + 513 pp, $129. [ISBN: 90-277-2605-1] Nature frequently comes to an equilibrium position that is also an optimal solution to some problem. The underlying theme of this book is to develop optimizing algorithms by trying to imitate nature's process of finding equilibria, an approach largely untried, which has yielded some promising results. Results in this book deal with the numerical solution of constrained optimization problems in classic mechanics and thermodynamics. Note the price! AWR

Dynamical Systems, P. *Dynamics of Infinite Dimensional Systems.* Ed: Shui-Nee Chow, Jack K. Hale. NATO ASI Series F, V. 37. Springer-Verlag, 1987, ix + 514 pp, $85. [ISBN: 0-387-18374-4] A representative sample of the presentations given at the 1986 NATO Advanced Study Institute on Dynamics of Infinite Dimensional Systems. The conference was an attempt to transmit ideas among investigators who are working with partial differential equations and functional differential equations as dynamical systems on function spaces. Includes forty papers from diverse areas (classical differential equations to application for infectious disease modelling). LB-E

Control Theory, T(16-17: 1). *Functional Analysis and Control Theory: Linear Systems.* Stefan Rolewicz. Math. & Its Applic. D Reidel (US Distr: Kluwer Academic), 1986, xv + 524 pp, $118. [ISBN: 90-277-2186-6] Functional analysis is discussed in a way which allows non-specialists to understand the applications which follow. The short course in functional analysis is followed by a development of the theory of general linear systems, application of that theory to ordinary and partial differential equations, and fundamentals of differential equations in Banach space. Examples come from physics and engineering. LB-E

Probability, T(18), P. *Stochastic Modelling and Analysis: A Computational Approach.* Henk C. Tijms. Ser. in Prob. & Math. Stat. Wiley, 1986, xii + 418 pp, $36.95. [ISBN: 0-471-90911-4] The main concern of this text is the application of stochastic models to practical situations. Some of the applications are renewal processes to inventory/production, Markov chains and operations research, and applications to queueing theory. MZ

Probability, S(13). *Probability Theory (First Steps).* E.S. Wentzel. Transl: N. Deineko. MIR (US Distr: Imported Pub), 1986, 87 pp, $1.95 (P). [ISBN: 0-8285-3196-X] Introduction to probability for non-mathematics majors. Includes probability as frequency, basic rules of probability, and random variables. Examples indicate wide applications. Brief, and easy to read. MS

Stochastic Processes, T(18: 1), P*. *Introduction to Stochastic Differential Equations.* Thomas C. Gard. Pure & Appl. Math., V. 114. Marcel Dekker, 1988, xi + 234 pp, $65. [ISBN: 0-8247-7776-X] Basic account, requiring only introductory courses in real analysis, ordinary differential equations, and some knowledge of continuous time stochastic processes. Contains review of needed probability theory. Emphasis on role of Ito's formula in stochastic analysis. Applications, including population dynamics and sample path approximations. Exercises; reference list. JK

Elementary Statistics, T(13: 1). *Statistics and Data Analysis, An Introduction.* Andrew F. Siegel. Wiley, 1988, xxv + 518 pp, $33. [ISBN: 0-471-87659-3] Introductory text with a multitude of examples. Objectives clearly defined in each section; end-of-chapter summary, questions, and problems. Exploratory data analysis incorporated throughout. Real data sets used illustrating a wide range of applications. MS

Statistics, P. *Statistical Decision Theory and Related Topics IV.* Ed: Shanti S. Gupta, James O. Berger. Springer-Verlag, 1988, $32 each. *Volume 1,* xii + 418 pp [ISBN: 0-387-96661-7]; *Volume 2,* xvi + 399 pp. [ISBN: 0-387-96662-5] 65 invited papers and discussion from a symposium and workshop held at Purdue University in June 1986. First volume devoted primarily to conditioning and Bayesian ideas; second volume to aspects of decision theory. LAS

Statistics, T(13: 1, 2), P. *Statistical Methods for Business and Economics, Third Edition.* Roger C. Pfaffenberger, James H. Patterson. Richard D Irwin, 1987, xiv + 1246 pp, $38.95. [ISBN: 0-256-03664-0] Material presented in non-mathematically rigorous fashion. Extensive use of examples. An increased number of problems. Minitab used throughout. SAS also used for regression. Knowledge of calculus is limited to advanced sections. (*Second Edition,* TR, February 1982.) MS

Statistics, P. *The Teaching of Practical Statistics.* C.W. Anderson, R.M. Loynes. Ser. in Prob. & Math. Stat. Wiley, 1987, xi + 199 pp, $52.95. [ISBN: 0-471-91572-6] Discussion of abilities for the ideal statistician; aims of statistics teaching; ways of teaching statistical practice. Describes use of sequence of projects tailored to different backgrounds and needs. Wonderful collection of project suggestions. MS

Statistics, S, P. *Lecture Notes in Statistics-43: Majorization and the Lorenz Order: A Brief Introduction.* Barry C. Arnold. Springer-Verlag, 1987, vi + 122 pp, $16.30 (P). [ISBN: 0-387-96592-0] Let R_n^+ be the positive orthant of R_n, and let $[x_{1:n}, x_{2:n}, \ldots, x_{n:n}]$ be the coordinates of $\mathbf{x} \in R_n^+$ arranged in decreasing order. Hardy, Littlewood, and Polya (as well as others) have discussed a partial ordering on R_n^+ defined by writing $\mathbf{x} \geq \mathbf{y}$ if and only if $\sum_{i=1}^k x_{i:n} \geq \sum_{i=1}^k y_{i:n}, k = 1, 2, \ldots, n-1$ and $\sum_{i=1}^n x_{i:n} = \sum_{i=1}^n y_{i:n}$. This ordering is the majorization referred to in the title, and if interpreted in terms of income distribution in a population of n individuals, Robin Hood type redistributions of wealth can be defined. These lead to the concept of Schur convexity, and our author is on his way to a nice introduction to a class of inequalities that is treated on a much more advanced level by Marshall and Olkin in *Inequalities: Theory of Majorization and Its Applications* (TR, May 1980). AWR

Statistics, T(14-15: 1, 2), P. *Modern Mathematical Statistics.* Edward J. Dudewicz, Satya N. Mishra. Ser. in Prob. & Math. Stat. Wiley, 1988, xix + 838 pp, $44.80. [ISBN: 0-471-81472-5] Could be used as a probability text (Chapters 1-6), or as a probability and mathematics statistics text. Integrates modern and classical statistics results. Variety of examples. Fortran programs included; SAS emphasized. Honors problems included. MS

Statistics, P. *Multivariate Statistical Modeling and Data Analysis.* Ed: H. Bozdogan, A.K. Gupta. Theory & Decision Library, Ser. B. D Reidel (US Distr: Kluwer Academic), 1986, ix + 189 pp, $54. [ISBN: 90-277-2592-6] Proceedings of an advanced symposium held in Virginia in May, 1986. FLW

Programming, T(13). *Using Turbo Basic.* Frederick E. Mosher, David I. Schneider. Osborne McGraw-Hill, 1988, xix + 457 pp, $19.95 (P). [ISBN: 0-07-881282-8] An introduction to programming using Turbo Basic, this book contains introductory material for beginning programmers and more in-depth sections on the Turbo Basic environment for experience. Rudimentary understanding of Basic is assumed. Many example programs are included. Other topics include graphics, sound, and mathematical, scientific, and business applications. Example output is provided in the graphics section. LB-E

Languages, S(16-17), P, L. *An APL Compiler.* Timothy Budd. Springer-Verlag, 1988, xi + 156 pp, $22.50 (P). [ISBN: 0-387-96643-9] Report of a project to develop a compiler for APL, which heretofore had been thought impossible due to weak variable typing and other very high-level features. The result, a prototype compiler written in C for UNIX machines (and available from the author), is far from production quality, but is sufficient to resolve most algorithmic and efficiency issues. LAS

Computer Systems, P. *Distributed Computing: Structure and Complexity.* H.L. Bodlaender. CWI Tract, V. 43. Math Centrum, 1987, vi + 294 pp, Dfl. 45.40 (P). [ISBN: 90-6196-327-3] Presentation of original research in some theoretical distributed computing topics. Most of the text is devoted to characterization and analysis of emulations, a notion of simulation for networks and processors, including complexity of finding uniform emulations. Also: algorithms for decentralized extrema-finding, protocols for distribution of records in rings of processors; deadlock-free packet switching networks. RB

Computer Systems, C, P, L*.** *Mathematica: A System for Doing Mathematics by Computer.* Stephen Wolfram. Addison-Wesley, 1988, xviii + 749 pp, $29.95 (P); $44.25. [ISBN: 0-201-19330-2; 0-201-19334-5] The "definitive" user guide to the latest symbolic computer package named—at Steve Jobs suggestion—"Mathematica." Written by MacArthur Fellow Wolfram, who earlier designed the package SMP, *Mathematica* is designed in two parts—a kernel which runs the same on any (sufficiently powerful) computer, and a front end which uses special features (windows, graphics) of particular hardware. Intended to "liberate the prose of mathematics from its grammar," *Mathematica* comes with built-in capability to do most of calculus and linear algebra, but not much more (e.g., not statistics or differential

equations). Well linked to other computer systems (e.g., UNIX, TeX, Fortran, Postscript) *Mathematica* currently runs on Macintosh as well as on most UNIX systems. This guide provides an extensive example-based survey of *Mathematica* behavior, together with a full reference guide and a thorough explanation of system design (including how to extend the system through external *Mathematica* macro packages). LAS

Computer Systems, S(17-18), P. *Lecture Notes in Computer Science-284: Embedded Systems.* Ed: A. Kündig, R.E. Bührer, J. Dähler. Springer-Verlag, 1987, v + 207 pp, $21.80 (P). [ISBN: 0-387-18581-X] These lectures are an attempt to bridge the gap between newer research results in system architecture and its description on one hand, and actual application of those results in developing real time/process control systems on the other. The architecture lectures are not included; instead the design aspect of embedded systems is addressed. The text is appropriate for students with previous experience in the area. LB-E

Computer Graphics, C. *Advanced Graphics in C: Programming and Techniques.* Nelson Johnson. Osborne McGraw-Hill, 1987, xvi + 670 pp, $22.95 (P). [ISBN: 0-07-881257-7] A thorough coverage of the essential elements needed to do graphics in C. Contains techniques and algorithms for graphics programming, graphics editing, hard-copy creation, user access, and communications. Focuses on IBM EGA graphics standard. Johnson develops a fairly complete set of graphic functions from the ground up—explaining algorithms and possible alternatives as he goes. Additional functions include support for serial and parallel I/O, pop-up menus, and graphics text. Appendix contains the complete C codes for a graphics system GRAPHIQ (also available on disk). Step-by-step approach also gives a valuable lesson in software system development—modularity. PS

Theory of Computation, P. *Algorithmic Information Theory.* Gregory J. Chaitin. Tracts in Theoret. Comput. Sci., V. 1. Cambridge U Pr, 1987, xi + 175 pp, $29.95. [ISBN: 0-521-34306-2] This entire monograph of 175 pages is directed at the proof of a single important theorem in computer science—the Gödel Incompleteness Theorem. The approach taken to prove this theorem is quite different from that of virtually all other mathematicians. The author uses an information-theoretic approach based on the size of computer programs. GMS

Theory of Computation, P. *Lecture Notes in Computer Science-280: Mathematical Models for the Semantics of Parallelism.* Ed: Marisa Venturini Zilli. Springer-Verlag, 1987, v + 231 pp, $20.60 (P). [ISBN: 0-387-18419-8] Papers presented at the Conference on the Semantics of Parallelism held in Rome, Italy in September 1986. The eight papers discuss advanced research work on parallel processing on distributed systems and on guaranteeing integrity, security, and consistency of multiple processor systems. GMS

Theory of Computation, P, L*. *Open Problems in Communication and Computation.* Ed: Thomas M. Cover, B. Gopinath. Springer-Verlag, 1987, viii + 236 pp, $25. [ISBN: 0-387-96621-8] Contributions of over 50 open problems in the theory of communication and computation, with a few solutions discovered by participants, from summer seminars in 1984-86 at Bell Communications Research and Stanford University. Highlight is FRACTAN, J.H. Conway's playful programming syntax for fractions which, started with short lists of simple fractions, can generate (a) all primes, (b) the digits of π, or (c) any Turing machine. LAS

Theory of Computation, T(17-18: 2), P. *Theories of Computational Complexity.* Cristian Calude. Annals of Disc. Math., V. 35. North-Holland (US Distr: Elsevier Science), 1988, xii + 487 pp, $97.50. [ISBN: 0-444-70356-X] A comprehensive treatment of the field of computational complexity. Divided into five parts, each of which address one of the five major divisions of the field—recursive function theory, Blum's complexity theory, Kolmogorov's complexity theory, LOF theory, and programming hierarchies. It treats the subject in a highly advanced form and is appropriate only for professionals in this specialized field. GMS

Theory of Computation, P. *Lecture Notes in Computer Science-287: Foundations of Software Technology and Theoretical Computer Science.* Ed: Kesav V. Nori. Springer-Verlag, 1987, ix + 540 pp, $40 (P). [ISBN: 0-387-18625-5] A collection of approximately 30 papers and invited talks presented at the 7th International Conference on Theoretical Computer Science held at Pune, India in December 1987. The papers address advanced research issues in automata theory, algorithm analysis, logic programming, parallel processing, and software theory. GMS

Computer Science, T(18), P. *Parallel Complexity Theory.* Ian Parberry. Wiley, 1987, 200 pp, $22.95 (P). [ISBN: 0-273-08783-5] Foundations of parallel complexity theory for advanced graduate students or researchers in theoretical computer science. Assumes familiarity with automata theory, formal languages, complexity and analysis of algorithms. LC

Computer Science, P. *Lecture Notes in Computer Science-278: Fundamentals of Computation Theory.* Ed: L. Budach, R.G. Bukharajev, O.B. Lupanov. Springer-Verlag, 1987, xiv + 505 pp, $40 (P). [ISBN: 0-387-18740-5] 110 papers selected from 195 presented at a conference held in Kazan, USSR, June 1987. Sections at the conference were on mathematical models of computation, synthesis and complexity of control systems, probabilistic computations, theoretical problems of programming, and problems of computed-assisted deductions. RWN

Computer Science, P. *Lecture Notes in Computer Science-290: Co-oP.* Tung X. Bui. Springer-Verlag, 1987, xiii + 250 pp, $24.50 (P). [ISBN: 0-387-18753-7] Study of decision support system (DSS) methods for multiple user domains which use computer-assisted group interaction, with the goals of a cooperative decision setting and distributed and demo-

cratic decision making. The Co-oP system, based on a process-oriented group decision methodology, is discussed. RM

Computer Science, P. *Lecture Notes in Computer Science-279: Graph Reduction.* Ed: Joseph H. Fasel, Robert M. Keller. Springer-Verlag, 1987, xvi + 450 pp, $34.60 (P). [ISBN: 0-387-18420-1] 1986 workshop on graph reduction (an elegant model for evaluation of functional programming languages) in which computation is viewed as a succession of transformations between graphs. Papers include discussion of models, architectures, resource control issues, arrays. RM

Applications, P, L. *Mathematics as a Service Subject.* Ed: A.G. Howson, *et al.* ICMI Study Series. Cambridge U Pr, 1988, iv + 92 pp, $39.50; $14.95 (P). [ISBN: 0-521-35395-5; 0-521-35703-9] Editorial overview and eight lead papers derived from an April 1987 international symposium in Udine, Italy, organized by the International Commission on Mathematical Instruction (ICMI). Major emphasis is on teaching university mathematics to engineering students and on the importance of such topics as geometry and discrete mathematics. Opening survey paper by the editors sketches a comprehensive agenda of issues; a companion volume of selected papers presented at the symposium will be published in the Springer CISM Series. LAS

Applications (Biological Science), T(15-16: 1, 2), L. *Mathematical Models in Biology.* Leah Edelstein-Keshet. Math. Ser. Random House, 1988, xvi + 586 pp, $32. [ISBN: 0-394-35507-5] Nicely integrated treatment of discrete, continuous, and multivariable models in biology, including phase plane ordinary differential equation methods, linear stability, bifurcations, partial differential equations applied to population, cellular, molecular biology, physiology, etc. Biological models, with mathematical tools developed in tandem, where mathematics serves as unifying concept via themes of steady states, stability, parameter variations. Minimal prerequisites, very nice examples, projects, references. RM

Applications (Electrical Engineering), T(16-17: 2). *Digital Transmission Systems and Networks.* Michael J. Miller, Syed V. Ahamed. Elect. Eng., Commun., & Signal Proc., V. 9 & 10. Computer Science Pr, 1987, $36.95 each. *Volume I: Principles,* xix + 274 pp, [ISBN: 0-88175-094-8]; *Volume II: Applications,* xviii + 366 pp. [ISBN: 0-88175-176-6] Two-volume text which treats the engineering aspects of integrated digital systems for telephone, TV, data, and other services. *Volume 1* discusses evolution of ideas and principles; *Volume 2* discusses implementation. RM

Applications (Engineering), T(17: 1, 2), S, P, L. *Vibrations in Mechanical Systems: Analytical Methods and Applications.* Maurice Roseau. Springer-Verlag, 1987, xiv + 515 pp, $69. [ISBN: 0-387-17950-X] Translated from the 1984 French edition. On free or induced vibrations in discrete or continuous mechanical structures. Linear, nonlinear, asymptotic theories via analysis of ordinary and partial differential equations. Some previously unpublished results. Superb selection of examples of applications. For graduate students and researchers in mechanics and mechanical engineering. Many references. No problems. JK

Applications (Engineering), P*. *Structural Reliability: Analysis and Prediction.* R.E. Melchers. Civil Engin. Halsted Pr, 1987, 400 pp, $89.95. [ISBN: 0-470-20873-2] Probability applied to structural reliability. Practical aspects and underlying theory. Unified discussion of structural systems reliability, including review of influence of time. Simulation methods, modern second-moment and transformation methods. Relationship of human error and human intervention to failure probabilities. Appendices with summary of probability theory and appropriate tables. Numerous references. JK

Applications (Engineering), T(16-17: 1, 2), L. *Engineering Analysis: A Vector Space Approach.* Robert J. Schilling, Hua Lee. Wiley, 1988, xiv + 337 pp, $44.95. [ISBN: 0-471-82760-6] Aimed at seniors or first year graduate students in engineering, and assuming only a traditional background in calculus, the emphasis is on linear spaces and operators. The first seven chapters are an introduction to linear algebra. Chapters 8 and 9 treat differential and difference systems, and the final chapter deals with orthogonal transforms, including Fourier, Hankel and Hilbert transforms with applications. Exercises, references, index. JS

Applications (Fluid Mechanics), P*. *The Mathematical Theory of Turbulence, Second Edition.* M.M. Stanišić. Universitext. Springer-Verlag, 1988, xviii + 501 pp, $45 (P). [ISBN: 0-387-96685-4] New fourth chapter on the dispersion of temperature in magnetohydrodynamic turbulence from the Lagrange-Euler point of view. New Chapter V on contemporary developments in theoretical and experimental turbulence. References updated to 1987. (*First Edition,* TR, October 1985.) JK

Reviewers

RJA: Richard J. Allen, St. Olaf; DFA: David F. Appleyard, Carleton; RB: Richard Brown, Carleton; LB-E: Lisa Brown-Ewing, Macalester; JNC: Judith N. Cederberg, St. Olaf; LC: Laura Chihara, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; RSF: Robert S. Fisk, St. Olaf; SG: Steven Galovich, Carleton; GG: George Gilbert, St. Olaf; BH: Bruce Hanson, St. Olaf; RH: Rhonda Hatcher, St. Olaf; RK: Roger Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; AM: Alan Magnuson, St. Olaf; SM: Steve McKelvey, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; MR: Matthew Richey, St. Olaf; AWR: A. Wayne Roberts, Macalester; GMS: G. Michael Schneider, Macalester; PS: Paul Schoening, St. Olaf; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MS: Myriam Steinback, Macalester; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; LW: Lynne Walling, St. Olaf; FLW: Frank L. Wolf, Carleton; MZ: Michael Zielinski, St. Olaf.

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 95, Number 9 November 1988

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Dept. of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Shorter articles about the teaching of mathematics should be sent to either MELVIN HENRIKSEN (linear algebra, algebra, differential equations), Department of Mathematics, Harvey Mudd College, Claremont, CA 91711, or STAN WAGON (calculus, analysis, number theory, discrete mathematics, statistics), Department of Mathematics, Smith College, Northampton, MA 01063. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to ANITA E. SOLOW, Dept. of Mathematics, Grinnell College, Grinnell, IA 50112; in geometry or topology to DENNIS DETURCK, Dept. of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to DAVID J. HALLENBECK, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Proposed problems (three copies) and solutions (two copies), both elementary and advanced to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, University of Calgary, Alberta, Canada T2N 1N4.

Reprint Permission: A. B. WILLCOX, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: MS. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $26 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $29. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $90 per year.

Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Assocation of America, 1529 Eighteenth Street, N.W., Washington, DC 20077-9564.

PRINTED IN THE UNITED STATES OF AMERICA

How I Became a Torchbearer for Matrix Theory*

OLGA TAUSSKY, *California Institute of Technology*

OLGA TAUSSKY is sufficiently well known to *Monthly* readers that further words here would be superfluous. Besides, this article is in itself biographical, as you will see [*Ed.*].

Some time ago in our public library I picked up a book, *The Best of All Possible Worlds*, by Peter F. Drucker, a professor of social science at Claremont Graduate School. In that book, Drucker writes about a pupil of Riemann who was to write his thesis on quaternions. Riemann had seen their importance to his own areas of study, and his student saw that they would lead to the subject that we now call matrix algebra, which has become all the rage.

But matrices were not always the rage. They have played a large role in group theory since the work of Elie Cartan, and they play a role in physics and in statistics. Still, matrix theory reached me only slowly. Since my main subject was number theory, I did not look for matrix theory. It somehow looked for me. In what follows a number of instances of such events are sketched.

1. Algebraic number theory. In a proof given by Minkowski in 1900** there appears the theory of matrices with dominant diagonal. He was reproving some of Dirichlet's results concerning units in algebraic number fields, and he observed that if a real matrix A has negative off-diagonal elements but only positive row sums, then its determinant is nonzero, in fact even positive. A similar theorem can be proved by a continuity argument. Such results are connected with the famous Geršgorin theorem, to which I will return in section 6 below.

2. The theorem of Shoda and R. C. Thompson. The next eye-opener for me was much stronger and it appeared in stranger circumstances: one of my favorite theorems in matrix theory came to me via class field theory! It is K. Shoda's theorem concerning matrices of determinant 1, and it states that for certain fields F, such a matrix is a commutator $B^{-1}C^{-1}BC$ where B, C are matrices over F.

How does class-field theory come into this? When I was working on my thesis in class-field theory, a subject created by Hilbert, some of the most important results in the subject were being found by Takagi, in Japan. I summoned up the courage to ask him for reprints, and not only did he send them, but a number of other mathematicians in Japan, even in quite different areas, sent reprints too, and one of

*This is a slightly enlarged version of an invited lecture delivered in Raleigh, North Carolina, at the 1985 SIAM Conference on Applied Linear Algebra, organized by R. Brualdi and H. Schneider.

**Bibliographic references are collected at the end of the paper, and are arranged by sections.

them was Shoda. I adored this theorem and felt right away that some day it would play a role in my life (but it took a long time to do so).

The problem was this. If we take a pair of *nonsingular* matrices X, Y over a field, then it is easy to see that they can be replaced by AB, BA, for suitable matrices A, B if and only if they are similar. I tried to find out under what circumstances there exist matrices A, B, C such that

$$X = ABC, \qquad Y = CBA.$$

The answer is, iff $\det X = \det Y$, or equivalently, iff $\det XY^{-1} = 1$. My proof depended on Shoda's theorem, so it was valid in fields for which that theorem holds. Shoda himself had not found all of those fields, but R. C. Thompson, in his 1960 thesis, was able to characterize them completely. The result is now called the Shoda-R. C. Thompson theorem (see also Sourour 1986, for recent work in this area).

3. Topological algebra. Pontrjagin's paper *Über stetige algebraische Körper* fell into my hands quite by accident. Let F be a topological field, i.e., a field that is also a topological space, so that

$$\lim(a_n + b_n) = \lim a_n + \lim b_n; \qquad \lim(a_n b_n) = \lim a_n \lim b_n;$$

$$\lim b_n^{-1} = (\lim b_n)^{-1} \qquad \text{unless} \lim b_n = 0.$$

Then Pontrjagin proves that under certain topological assumptions, F must be isomorphic with either the reals, the complex numbers or the real quaternions. Under the same conditions, F must contain a subfield D isomorphic with the reals, such that D commutes with every element of F. Furthermore, F contains a finite set of elements such that every element of F is a linear combination of these elements with coefficients in D. Hence F is actually a division algebra over the reals, and so by a theorem of Frobenius it is isomorphic with the reals, the complex numbers or the real quaternions. Frobenius used matrix theory for the proof of his theorem. Later I myself found a topological proof for the theorem. Pontrjagin's proof partitioned the field into sets $\{\lambda\}, \{\mu\}, \{\nu\}$ with the properties that $\lambda^n \to 0$, μ^n is divergent, and ν^n has no divergent subsequence nor is 0 a point of accumulation (analogous to the interior, exterior and circumference, respectively, of the unit disk).

I became interested in studying such λ sets for real algebras. The matrices over the reals form such a set in their natural topology. John Todd and I wrote the paper 'Infinite powers of matrices' on this subject.

It is well known that matrices C whose powers approach zero play a big role in iteration processes. They have been characterized by P. Stein as being precisely those matrices C for which there exists a matrix X such that $X - CXC^*$ is positive definite. Later I showed that Stein's theorem is equivalent, by a Cayley transformation, to the Lyapunov theorem for stable matrices. See also section 6 below for more on Lyapunov's theorem.

4. Integral matrices. An integral matrix is one whose entries are rational integers. My first work on this subject was with John Todd, during the last weeks before the war broke out, in Great Britain. Integral matrices had made me return to my major subject of number theory in a big and unexpected way. It happened at Bryn Mawr College, where I held a scholarship during a year when Emmy Noether was also there, and she was known as a champion of completely abstract approaches, even to

number theory. But a strange thing happened. Another Fellow at the College, Grace Shover (now Quinn), introduced me to her thesis adviser MacDuffee, who was an expert in matrices. I learned more about MacDuffee's work, some of which can be traced back to Poincaré's studies on matrices that are attached to ideals in algebraic number fields. This led to my work on so-called ideal matrices. Another paper that was influential for me was that of Latimer and MacDuffee, because it provides an important link between algebraic number theory and integral matrices.

5. The theorem of McCoy. John Todd and I spent the first year of World War II, 1939–40, in Ireland on leave from our positions in the University of London. I had no assigned duties during the first term, and found the library at Queen's University very appealing. Among items that I had not seen before was the work of McCoy, in particular his well known characterization of pairs of matrices which can be transformed simultaneously to upper triangular form.

Precisely, here is McCoy's result. Let A, B be two $n \times n$ matrices, and suppose they have the following property: for every polynomial f in two variables, the matrix $f(A, B)$ has for its eigenvalues the numbers $f(\lambda, \mu)$, where λ, μ are suitably ordered eigenvalues of A and B, respectively. McCoy showed that for A, B to have this property it is necessary and sufficient that a matrix T exists such that both $T^{-1}AT$ and $T^{-1}BT$ are upper triangular.

Since this work was kind enough to jump off of the library shelves into my hands, I started a correspondence with McCoy, and my interest in matrices was further strengthened. Drazin, Dungey, and Gruenberg [1951] gave a more elementary proof of McCoy's theorem, but my 1957 proof, via the radical in abstract algebras, takes only a few lines. See also Flanders' treatment.

6. The new numerical mathematics. In the early forties, numerical mathematics was in a rather primitive state. The standard textbook in Britain was Whittaker and Robinson, which concentrated on interpolation, the solution of equations by Newton's method, numerical quadrature and Fourier analysis. The solution of linear equations was studied, not for its own sake, but as an appendage of the theory of least squares approximation. Characteristic value problems and differential equations received little attention. A book with a more comprehensive point of view was that of Frazer, Duncan and Collar, but it was not as widely used.

In the decades that followed there was intensive study of the approximate solution of continuous problems such as differential equations, and many of these involved matrix methods in a natural way, either iteratively or directly. In particular, the whole subject of sparse matrices grew largely out of the study of characteristic value problems and the discretization of partial differential equations. I had several contacts with these developments, which I will now discuss. They brought me into matrix theory again, whereas previously, numerical mathematics had interested me only for special problems in number theory.

(A) *The Geršgorin theorem*

The war was on, WWII, and I was working in London at the National Physical Laboratory under R. A. Frazer in the flutter group. I was assigned to the study of flutter in supersonic aircraft, which leads to boundary value problems in hyperbolic partial differential equations. Hence this work did not immediately contribute to my matrix enthusiasm. However, I had read Frazer's article on how the flutter calcula-

tions were to be carried out. A large group of young girls, drafted into war work, did the calculation on hand-operated machines, following the instructions of Frazer and his assistants.

The relevance of these calculations to aircraft design is that in flight the interaction between the elastic forces in the airframe and the aerodynamic forces induces self-excited vibration which, above a certain speed, is unstable. This phenomenon is called flutter. It is, therefore, important to know what the flutter speed is before the aircraft is built and flown.

By a mere accident I had heard about the Geršgorin theorem, whose statement is given in a *Zentralblatt* review. It showed me how to reduce the amount of calculation, in a way that I will now try to explain.

The theorem itself states that the eigenvalues of an $n \times n$ matrix A with complex entries lie in the union of the closed disks ('Geršgorin disks')

$$|z - a_{ii}| \leqslant \sum_{k \neq i} |a_{ik}| \qquad (i = 1, 2, \ldots, n)$$

in the complex z plane. We will call the union of these disks the *Geršgorin set* of the matrix A, and will denote it by Γ.

In the case that I was working on, the question came down to showing that a certain 6×6 matrix of the form $-\omega^2 A + i\omega B + C$, where ω, the flutter parameter, is taken as 1 in the example (see Fig. 1), had no real eigenvalue to the left of the small circles. The matrix entries were in the neighborhood of 20 or so, but in itself that told us nothing about the whereabouts of the eigenvalues.

However, it turned out that the Geršgorin disks looked like the ones shown in FIG. 1.

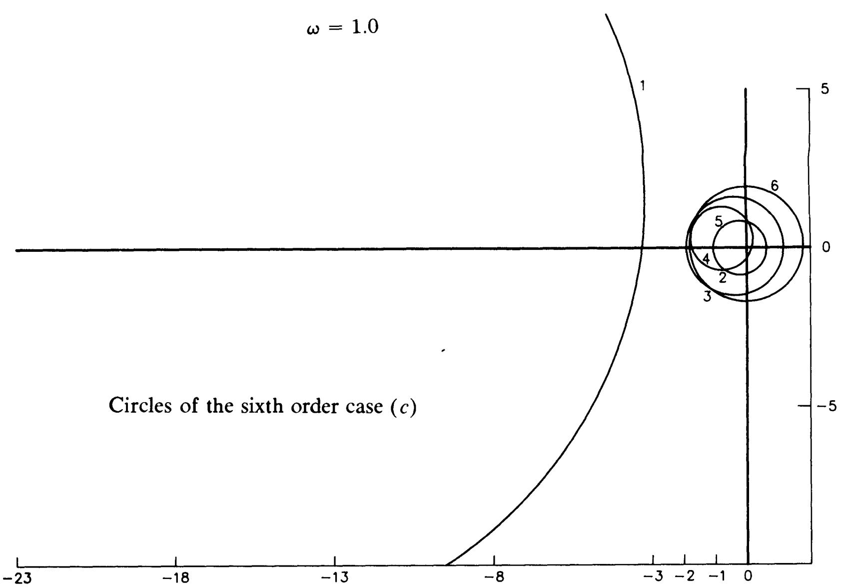


FIG. 1. The Geršgorin disks.

This meant that we were lucky indeed, because the Geršgorin theorem can be applied as follows:

(i) If Γ, the Geršgorin set, falls into two connected components, one generated by r of the disks and the other by s of them, then there are r eigenvalues in the first component and s in the other.

(ii) A similarity transformation $S^{-1}AS$ does not change the eigenvalues of A, but it may well change the estimates that are given by Geršgorin's theorem. Hence, by a careful choice of S, we may get sharper estimates. The easiest kind of an S to use, it turns out, is one that agrees with the identity matrix except in one position, say $S_{ii} = r$, where $r \neq 0$. This similarity will multiply the radius of the ith Geršgorin disk by $1/r$ while leaving all the centers of the disks unchanged.

(iii) The intersection of two Geršgorin sets, the original one and the one obtained after the similarity transformation, is again a region in which all of the eigenvalues must lie.

In FIG. 2, for example, it turned out that the larger circle could be replaced by the small circle far above the x-axis, by a similarity transformation of the diagonal type discussed above.

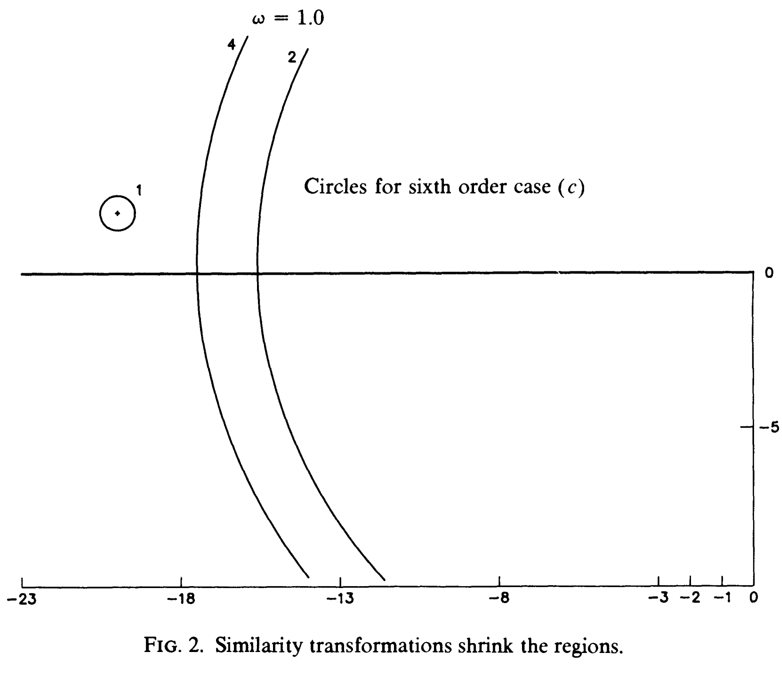


FIG. 2. Similarity transformations shrink the regions.

The small circle cannot be shrunk by the same method into a point unless the diagonal element is itself an eigenvalue. If we vary the diagonal similarities then at a certain point the other circles will overlap the isolated one and it will no longer be isolated. I raised the question of describing when this happens. Henrici, with a followup by F. Gaines, had the first contribution, in a special case. John Todd and Richard Varga gave different solutions of this problem. The school of A. Brauer at Chapel Hill worked on related questions.

Once again, I didn't ask to be assigned to matrix problems. They found me.

(B) *The Stein-Rosenberg theorem*

Mordell asked me to look at a manuscript of Stein and Rosenberg, before it went out to a referee. I liked it, and it has since become a classic. It has been particularly studied by Francois Robert in France, and is discussed in detail in Varga's 1962 book. The Perron-Frobenius theorem, which concerns the eigenvalues of matrices with nonnegative entries, plays a big part in this paper.

The Stein-Rosenberg theorem itself concerns iteration. The two classical iterative methods for the solution of linear equations $Ax = b$ can be described as follows. Assume $A = I - L - U$, where L, U are lower and upper strictly triangular and I is the unit matrix. Make a guess $x^{(0)}$. Then, in the Jacobi method, we improve the guess by means of

$$x^{(1)} = -(L + U)x^{(0)} + b.$$

Anyone who tries this soon discovers that it works better if one replaces the components of $x^{(0)}$ as fast as the improved ones are calculated, instead of continuing to use the old values until all new ones have been found. Formally, if we do that, the iteration process is defined by

$$x^{(1)} = (I - L)^{-1}(Ux^{(0)} + b)$$

and the procedure is called the Gauss-Seidel method. It seems plausible to assume that the Gauss-Seidel method, since it uses improved estimates of the components of the unknown vector, would converge faster than the Jacobi method. This is not the case always. However, the Stein-Rosenberg theorem asserts that if L and U are nonnegative, the two processes converge or diverge together, and the Gauss-Seidel process is at least as fast, whichever happens.

(C) *The Lyapunov theorem*

The Lyapunov theorem, too, is of great interest in flutter work. It is a criterion for the stability of an $n \times n$ complex matrix, where a stable matrix is one all of whose eigenvalues have negative real parts. The economist Arrow had used this theorem for a measure of the stability of an economic system.

The Lyapunov theorem gives the following criterion for a complex matrix A to be stable: there should exist a positive definite matrix H such that $AH + HA^* = -I$. This criterion is useful for theoretical purposes though it is not well suited to computation in particular instances.

(D) *The Hilbert matrix*

The Hilbert matrix A is defined by

$$a_{i,j} = 1/(i + j) \qquad (i, j = 1, 2, \ldots, n).$$

Its inverse is integral.

This matrix, too, made a surprise entry into my research interests. In late 1947, after we had settled down at the U.S. National Bureau of Standards, I received a letter from Professor G. Temple in London. He wrote that the Oscillation Subcommittee of the British Aeronautical Research Council was interested in the Hilbert matrix, and would appreciate comments from me.

In due course I wrote a paper explaining the slow convergence of the largest eigenvalue λ_n of A to its limiting value of π. John Todd then studied the condition of A and he and others used it as an example of an ill-conditioned matrix. Its

determinant, for example, is nonzero, but extremely small. Later H. S. Wilf and N. G. de Bruijn described the behavior of λ_n more precisely. While I had proved that

$$\lambda_n = \pi\{1 + O(1/\log n)\}$$

they showed

$$\lambda_n = \pi - \tfrac{1}{2}\pi^5(\log n)^{-2} + O\big((\log\log n)(\log n)^{-3}\big).$$

The literature on A is extensive and interest in A continues (see Wilf, *Ergebnisse*, volume 52, 1970).

7. The Perron-Frobenius theorem and combinatorial matrix theory. The Perron-Frobenius theorem concerns the eigenvalues of matrices that are irreducible and have nonnegative entries. Among its conclusions, for instance, is the fact that the matrix must have a positive real eigenvalue that is not exceeded, in absolute value, by any other eigenvalue.

It received a special lift through the proof by Wielandt (1950), cf. Gantmacher, Matrix Theory II. Graph theory plays a role there, in fact one can define irreducibility of a matrix by connectedness of a certain graph that is determined by the positions of the nonzero matrix entries.

I studied $N \times N$ incidence matrices A of projective planes, and observed that A^{N-1} always has strictly positive entries. I further raised the question of determining the exponent of A, i.e., the least power of A that has positive entries only. A. L. Dulmage and N. S. Mendelsohn showed that $A^4 > 0$ and that permutation matrices P, Q exist such that $(PAQ)^3$ has some zero entries. These authors used purely combinatorial arguments rather than the Perron-Frobenius theorem. For other combinatorial matrix work I refer to, e.g., Brualdi, Schneider, Engel, M. Hall, Ryser, etc.

8. Connections with topological algebra. In the late thirties I suddenly realized that the Cauchy-Riemann equations and the fact that they imply the Laplace equation can be expressed via matrix theory and can be connected with the fact that the field of complex numbers has no zero divisors. I then studied the values of n for which generalized 'Cauchy-Riemann' equations for n functions u_i in n real variables x_i lead to the n-dimensional Laplace equation. I used algebras over the reals which have no zero divisors.

At that time the fact that such algebras must have dimensions $1, 2, 4, 8$ was not yet completely established, although great progress had been made by topological methods, in particular by E. Stiefel. However, Stiefel reproved my result about dimensions $1, 2, 4, 8$ and for complex variables as well, without topology, by using representation theory of algebras and matrix theory.

9. M. Marden's book on geometry of polynomials. Marden's book, which appeared as *Mathematical Surveys*, No. 3, AMS, 1966, especially section 31, contains applications of matrix theory to the study of geometry of the zeros of polynomials, e.g. Geršgorin's theorem applied to the companion matrix of a polynomial gives information on the location of its zeros, and the theorem of Perron-Frobenius does likewise (cf. H. S. Wilf (1961)).

10. The Schur matrix. The Schur matrix is the matrix $(e^{2\pi imn/q})$ for $1 \leqslant m, n \leqslant q$, and it has generated many problems of great interest. Its trace, for example, is obviously a Gaussian sum. The eigenvalues of this matrix were obtained by Carlitz, and the eigenvectors by P. Morton answering a question of Hugh Montgomery. Landau's famous book on number theory had introduced me to this matrix, and it is a valuable bridge between matrices and the theory of numbers.

If we take the Schur matrix in the form

$$S = S_n = \left(\zeta^{(i-1)(j-1)}\right)_{i,j=1}^{n} \qquad (\zeta = \exp 2\pi i/n)$$

then we find it also in the theory of the Fast Fourier Transform. If $\mathbf{v}$ is a vector, then $S\mathbf{v}$ is essentially its discrete Fourier transform. Theilheimer showed that the Fast Fourier Transform, which speeds up the calculation quite a bit, amounts essentially to a factorization of S, namely,

$$S_{2n} = P\begin{bmatrix} T & 0 \\ 0 & T \end{bmatrix}\begin{bmatrix} I & I \\ D & \zeta^n D \end{bmatrix}$$

where P is a permutation matrix, $T = (\zeta^{2(i-1)(j-1)})$, and $D = \operatorname{diag}(1, \zeta, \dots, \zeta^{n-1})$.

11. Cramped matrices. A matrix theorem that fascinated me at a very early time came from the book by Speiser on finite group theory. It is connected with the fact that every finite group is isomorphic to a group of unitary matrices. Let A, B be such matrices, and put $C = ABA^{-1}B^{-1}$. Let the eigenvalues of A lie on an arc less than a semicircle of the unit circle. If A and C commute then $C = I$.

Zassenhaus wrote a paper, as did M. Marcus and R. C. Thompson, on this subject, and so had I done. Papers in functional analysis were stimulated too, and Berberian introduced the term 'cramped' for such matrices A. The theorem itself goes back to C. Jordan.

12. Other areas. While I was still in Britain, the idea had come to me that while there are a number of inequalities tying the eigenvalues of A, B and $A + B$ there had not been much done on explicit relationships between them. With these questions in mind I became very enthusiastic about Mark Kac introducing me to what I later called the L-property, L for linear, because the eigenvalues of $\lambda A + \mu B$ are then $\lambda\alpha_i + \mu\beta_i$ with α_i, β_i suitably ordered eigenvalues of A, B. Motzkin joined me in this work, and we proved among other things the following theorem: if for all λ and μ the matrix $\lambda A + \mu B$ is diagonalizable, then A and B commute. This result was reproved by Tosio Kato, using perturbation theory, and was later generalized by S. Friedland. Kaplansky generalized the L-property to operators.

After the war, in 1947, we came to the USA when John Todd was invited to work at the National Bureau of Standards on the uses of high speed computing machines, and soon after my arrival I was given employment as well. At that time I picked up the torch. Matrix theory had become a subject for me. Matrices were not any longer just used, they were algebraic structures like rings, groups, lattices....

At first I collected relevant references and they went into my chapter in John Todd's *Survey of Numerical Analysis*. But I was also working on other projects

linked to matrices, both algebraically and arithmetically. Stimulated by McCoy's work, I became interested in pencils of matrices and in matrix algebras. The Caltech theses by F. Gaines and by H. Shapiro generalized McCoy's idea of simultaneous triangularization to simultaneous block triangularization, and further contributions were made by R. C. Thompson. All of this led to the study of generalized commutativity, commutators, higher order commutators, and to my idea of polynomials in the commutator operator.

While still at NBS, in 1951, I was asked to organize a conference in numerical analysis and, of course, I chose matrix theory as the theme. It was perhaps the first matrix conference ever. The *Proceedings* of this conference contain Givens' rotation method for determining the eigenvalues and eigenvectors of real, symmetric matrices.

In my position at NBS I encouraged people to turn to matrix theory, and I thought of bounds for eigenvalues as my major project there. The school of W. V. Parker at Auburn also worked on this subject at the time. I had a master's student, Marion Walter, at NYU, who wrote a thesis on limits for the characteristic roots of a matrix, and I gave a course on matrix theory at NYU in 1955. I also looked after a group of highly talented postgraduate students and high ranking visitors at NBS. Strangely enough, they all became interested in matrix theory.

Matrix theory changed my life quite a lot. During the war and my civil service work I had lost my favorite subject to a large extent, though not entirely. However, integral matrices brought it back to me quite unexpectedly.

During our 1947 visit to Princeton I met Chowla. At that time he and two other mathematicians talked to me about similarity classes of integral matrices with the same characteristic polynomial. This reminded me of the theorem of Latimer and MacDuffee; Chowla urged me to find a new proof for it, and I managed to do so.

Since then integral matrices have played a major role for me, and I helped to get a number of other people, like Zassenhaus, Dade, M. Newman, Estes, and Guralnick interested in them. We found simpler ways to prove many classical number-theoretic theorems by methods that used integral matrices. I have given an exposition of the connection between algebraic number theory and integral matrices in an appendix to a book by H. Cohn on algebraic number theory.

Some advice. When you observe an interesting property of numbers, ask if perhaps you are not seeing, in the 1×1 case, an interesting property of matrices. Think of $GL(n, F)$ or $SL(n, F)$, $GL(n, Z)$ or $SL(n, Z)$.

When you have a pair of interesting matrices study the pencil that they generate, or even the algebra.

When the determinant of a certain matrix turns out to be important, ask about the matrix as a whole, for instance as in the case of the discriminant matrix, as suggested by the discriminant of an algebraic number field.

When a polynomial in one variable interests you, ask about the matrices of which it is the characteristic polynomial.

When people look down on matrices, remind them of great mathematicians such as Frobenius, Schur, C. L. Siegel, Ostrowski, Motzkin, Kac, etc., who made important contributions to the subject.

I am proud to have been a torchbearer for matrix theory, and I am happy to see that there are many others to whom the torch can be passed.

BIBLIOGRAPHY

Section 1

1. J. L. Brenner, A bound for a determinant with dominant main diagonal, *Proc. Amer. Math. Soc.*, 5 (1954) 631–634.
2. ______, Neuer Beweis einer Satzes von Taussky and Geiringer, *Arch. Math.*, 7 (1956) 274–275.
3. S. Geršgorin, Über die Abgrenzung der Eigenwerte einer Matrix, *Izv. Akad. Nauk S.S.S.R.* (1931) 749–754.
4. H. Minkowski, Zur Theorie der Einheiten in den algebraischen Zahlkörpern, *Gött. Nachr.* (1900) 90–93.
5. A. Ostrowski, Über die Determinanten mit überwiegender Hauptdiagonale, *Comm. Math. Helv.*, 10 (1937) 69–96.
6. G. B. Price, Bounds for determinants with dominant principal diagonal, *Proc. Amer. Math. Soc.*, 2 (1951) 497–502.
7. H. Schneider, An inequality for latent roots of a matrix applied to determinants with dominant main diagonal, *J. London Math. Soc.*, 28 (1953) 8–20.
8. O. Taussky, A recurring theorem on determinants, *Amer. Math. Monthly*, 56 (1949) 673–676.

Section 2

1. J. L. Brenner and J. S. Lim, The matrix equations $A = XYZ$ and $B = ZYX$ and related ones, *Bull. Amer. Math. Soc.*, 17 (1974) 179–183.
2. D. R. Estes, Scalar matrices as multiplicative commutators having prescribed determinants for the variables, *Linear and Multilinear Algebra*, 8 (1980) 213–217.
3. K. Fan, Some remarks on commutators of matrices, *Arch. Math.*, 5 (1954) 102–107.
4. H. Flanders, Elementary divisors of AB and BA, *Proc. Amer. Math. Soc.*, 2 (1951) 871–874.
5. K. Shoda, Einige Sätze über Matrizen, *Jap. J. Math.*, 13 (1937) 361–365.
6. O. Taussky, Generalized commutators of matrices and permutations of factors in a product of three matrices, Studies in Mathematics and Mechanics presented to Richard von Mises, Academic Press, NY, 1954.
7. R. C. Thompson, Commutators in the special and general linear groups, *Trans. Amer. Math. Soc.*, 101 (1961) 16–33.
8. ______, On matrix commutators, *Portugal. Math.*, 21 (1962) 143–153.
9. ______, Commutators of matrices with prescribed determinant, *Canad. J. Math.*, 20 (1968) 203–221.
10. ______, Commutators of matrices with coefficients from the field of two elements, *Duke Math. J.*, 29 (1962) 367–373.

Section 3

1. L. Pontrjagin, Über stetige algebraische Körper, *Ann. Math.*, 33 (1932) 163–174.
2. O. Taussky and John Todd, Infinite powers of matrices, *J. London Math. Soc.*, 17 (1942) 147–151.
3. O. Taussky, Matrices C with $C^n \to 0$, *J. Algebra*, 1 (1954) 5–10.

Section 4

1. O. Taussky and John Todd, Matrices with finite period, *Proc. Edinburgh Math. Soc.*, 6 (1939) 128–134.
2. O. Taussky, On a theorem of Latimer and MacDuffee, *Canad. J. Math.*, 1 (1949) 300–302.
3. ______, Ideal matrices, I., *Archiv. d. Math.*, 13 (1962) 275–282.

Section 5

1. M. P. Drazin, J. W. Dungey, and K. W. Gruenberg, Some theorems on commutative matrices, *J. London Math. Soc.*, 26 (1951) 221–228.
2. H. Flanders, Methods of proof in linear algebra, *Amer. Math. Monthly*, 63 (1956) 1–15.
3. R. M. Guralnick, Triangularization of sets of matrices, *Linear and Multilinear Alg.*, 9 (1980) 133–140.
4. T. Laffey, Simultaneous triangularization of matrices, *J. Algebra*, 44 (1977) 351–357.
5. N. H. McCoy, On the characteristic roots of matrix polynomials, *Bull. Amer. Math. Soc.*, 42 (1936) 592–600.

6. O. Taussky, Commutativity in finite matrices, *Amer. Math. Monthly*, 64 (1957) 229–235.
7. ______, Sets of complex matrices which can be transformed to triangular forms, *Coll. Math. Soc. János Bolyai*, 22 (1977) 579–590.

Section 6

1. J. L. Brenner, Geršgorin theorems by Householder's proof, *Bull. Amer. Math. Soc.*, 74 (1968) 625–627.
2. N. G. de Bruijn and H. S. Wilf, On Hilbert's inequality in n dimensions, *Bull. Amer. Math. Soc.*, 68 (1962) 70–73.
3. D. Carlson and H. Schneider, Inertia theorems for matrices, the semidefinite case, *Bull. Amer. Math. Soc.*, 68 (1962) 479–483.
4. M. Fiedler, Matrix inequalities, *Numer. Math.*, 9 (1966) 109–119.
5. R. A. Frazer, W. J. Duncan, and A. R. Collar, Elementary Matrices and Some Applications to Dynamics and Differential Equations, Cambridge University Press, 1938.
6. W. Givens, Elementary divisors and some properties of the Lyapunov mapping $X \to AX + XA^*$, Argonne National Laboratory ANL-6456.
7. ______, Numerical computation of the characteristic values of a real symmetric matrix, Oak Ridge National Laboratory, ORNL-1574, 1954.
8. A. J. Hoffman and H. Wielandt, The variation of the spectrum of a normal matrix, *Duke Math. J.*, (1953) 37–39.
9. M. A. Lyapunov, Problème général de la stabilité du mouvement, *Ann. Math. Studies*, 17, Princeton, 1949.
10. V. B. Lidskii, On the characteristic roots of a sum and a product of symmetric matrices, *Dokl. Akad. Nauk. SSSR*, 75 (1950) 769–772.
11. F. Robert, Autour du theórème de Stein-Rosenberg, *Numer. Math.*, 27 (1976) 133–141.
12. H. Schneider and A. Ostrowski, Some theorems on the inertia of general matrices, *J. Math. Anal. and Appl.*, 4 (1962) 72–84.
13. P. Stein and R. L. Rosenberg, On the solution of linear simultaneous equations by iteration, *J. London Math. Soc.*, 23 (1948) 111–118.
14. O. Taussky, A remark concerning the characteristic roots of finite segments of the Hilbert matrix, *Quart. J. Math. Oxford*, 20 (1949) 82–83.
15. O. Taussky, A remark on a theorem of Lyapunov, *J. Math. Anal. and Appl.*, 2 (1961) 105–107.
16. O. Taussky, A method for obtaining bounds for characteristic roots of matrices with applications to flutter calculations, Aeron. Res. Council of Great Britain, Report 10.508 (1947).
17. O. Taussky, Stable matrices, Programmation en mathématiques numériques, CNRS, No. 165, Besançon (1968) 75–88.
18. John Todd, On smallest isolated Gerschgorin disks for eigenvalues, *Numer. Math.*, 7 (1965) 171–175.
19. John Todd, Survey of Numerical Analysis, McGraw-Hill, New York, 1962.
20. John Todd, Computational problems concerning the Hilbert matrix, *J. Res. Nat. Bur. Standards*, 65 (1961) 19–22.
21. R. S. Varga, Matrix Iterative Analysis, Prentice Hall, 1962.
22. R. S. Varga, On smallest isolated Gerschgorin disks for eigenvalues, *Numer. Math.*, 6 (1964) 366–376.
23. H. Wielandt, An extremum property of sums of eigenvalues, *Proc. Amer. Math. Soc.*, 6 (1955) 106–110.
24. H. S. Wilf, Finite sections of some classical inequalities, *Ergeb. Mathematik*, 52 (1970).
25. H. S. Wilf, On finite sections of the classical inequalities, *Nederl. Akad. Wetensch. = Indag. Math.*, 24 (1962) 340–342.

Section 7

1. R. A. Brualdi, S. V. Parter, and H. Schneider, The diagonal equivalence of a nonnegative matrix to a stochastic matrix, *J. Math. Anal. Appl.*, 16 (1966) 31–50.
2. A. L. Dulmage and N. S. Mendelsohn, The exponents of incidence matrices, *Duke Math. J.*, 31 (1964) 575–584.
3. G. M. Engel and H. Schneider, The Hadamard-Fisher inequality in a class of matrices defined by eigenvalue monotonicity, *Linear and Multilinear Algebra*, 4 (1976) 155–176.

4. M. Fiedler and V. Pták, On matrices with non-positive off-diagonal elements and positive principal minors, *Czechoslovak Math. J.*, 12 (1962) 382–400.
5. F. R. Gantmacher, The Theory of Matrices, vol. II, 1959, translated by K. A. Hirsch, Chelsea, New York, 1960.
6. M. Hall, Finite projective planes, *Amer. Math. Monthly*, 62 no. 7, Part II (1955) 18–24.
7. J. C. Holladay and R. S. Varga, On powers of non-negative matrices, *Proc. Amer. Math. Soc.*, 9 (1958) 631–634.
8. V. Pták, On a combinatorial theorem and its application to non-negative matrices, *Czechoslovak Math. J.*, 83 (1958) 487–495.
9. H. Ryser, Geometries and incidence matrices, *Amer. Math. Monthly*, 62 no. 7, Part II (1955) 25–31.
10. H. Wielandt, Unzerlegbare nicht negative Matrizen, *Math. Z.*, 52 (1950) 642–648.

Section 8

1. E. Stiefel, Über Richtungsfelder in den projektiven Räumen und einen Satz aus der reellen Algebra, *Comment. Math. Helv.*, 13 (1941) 209–239 (Satz IIIc).
2. ______, On Cauchy-Riemann equations in higher dimensions, *J. Res. Nat. Bur. Standards*, 48 (1952) 395–398.

Section 9

1. M. Marden, Geometry of the Zeros of Polynomials, Math. Surveys 3, Amer. Math. Soc. 1966.
2. H. S. Wilf, Perron-Frobenius theory and the zeros of polynomials, *Proc. Amer. Math. Soc.*, 12 (1961) 247–250.

Section 10

1. L. Carlitz, Some cyclotomic matrices, *Acta Arithm.*, 5 (1959) 293–308.
2. P. Morton, On the eigenvectors of Schur's matrix, *J. Number Theory*, 12 (1980) 122–127.
3. I. Schur, Über die Gausschen Summen, *Nachr. Kgl. Ges. Göttingen Math.* (1921) 147–153.
4. F. Theilheimer, A matrix version of the Fast Fourier Transform, *IEEE Trans.* AU-17 (1969) 158–161.

Section 11

1. S. K. Berberian, A note on operators unitarily equivalent to their adjoints, *J. London Math. Soc.*, 37 (1962) 403–404.
2. C. Jordan, Mémoire sur les équations différentielles linéaires à intégrale algébrique, *J. für die Reine u. Angew. Math.*, 84 (1878) 89–215.
3. M. Marcus and R. C. Thompson, On a classical commutator result, *J. Math. Mech.*, 16 (1966) 583–588.
4. C. R. Putnam, Commutation properties of Hilbert space operators and related topics, *Ergeb. Math.*, 36 (1967).
5. O. Taussky, Commutators of unitary matrices which commute with one factor, *J. Math. Mech.*, 10 (1961) 175–178.
6. H. Zassenhaus, On a paper by O. Taussky, *J. Math. Mech.*, 10 (1961) 179–180.

Area-minimizing Surfaces, Faces of Grassmannians, and Calibrations

FRANK MORGAN, *Williams College*

FRANK MORGAN: I went to MIT and Princeton, where my thesis advisor, Fred Almgren, introduced me to minimal surfaces. Since then I have taught at MIT, Rice, and Stanford. In the fall of 1987, I began teaching at Williams College. (Photo courtesy of Gordon Graham/Prism.)

1. Faces of Grassmannians

A one-dimensional curve that heads in a constant direction (i.e., a straight line segment) is length-minimizing among all curves with the same boundary.

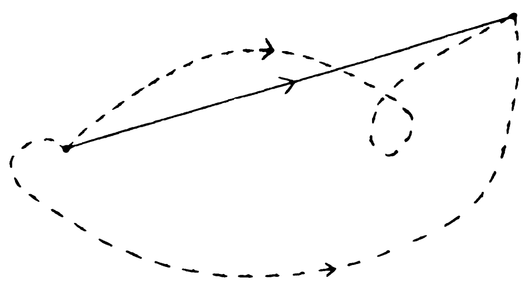


FIG. 1(1). A straight line is length-minimizing.

This fact has an interesting generalization to higher dimensions. Instead of curves, one considers m-dimensional surfaces S in n-dimensional space $\mathbb{R}^n$. Instead of a constant direction, one considers a certain collection of tangent plane directions. If all the tangent planes to a surface S belong to the same collection, then S is area-minimizing among all oriented surfaces with the same boundary.

These collections of tangent plane directions are called "faces." The set of all oriented m-dimensional planes through the origin in $\mathbb{R}^n$ is called the *Grassmannian* $G(m, \mathbb{R}^n)$. The definition of these faces depends on viewing the Grassmannian $G(m, \mathbb{R}^n)$ as sitting inside a higher-dimensional Euclidean space $\mathbb{R}^N$. The dimension N equals $\binom{n}{m} = n!/m!(n-m)!$, the number of m-dimensional axis planes in $\mathbb{R}^n$. Indeed, these axis planes give an orthonormal basis for $\mathbb{R}^{\binom{n}{m}}$. The $\binom{n}{m}$ coordinates of any given m-dimensional plane $\xi \in G(m, \mathbb{R}^n)$ are just the areas of the projections of a unit area in ξ onto the $\binom{n}{m}$ axis planes. Each coordinate must be signed $\pm$ according to whether ξ and the axis plane have similar or opposite orientations. It turns out that the sum of the squares of these coordinates is always 1, so that $G(m, \mathbb{R}^n)$ sits in the surface of the unit sphere in $\mathbb{R}^{\binom{n}{m}}$.

The shape of the Grassmannian $G(m, \mathbb{R}^n) \subset \mathbb{R}^{\binom{n}{m}}$ is independent of the original choice of orthonormal basis for $\mathbb{R}^n$, which determines the $\binom{n}{m}$ axis planes.

Now bring a hyperplane in $\mathbb{R}^N$ in from infinity until it first touches the Grassmannian. Usually it will hit at a single point, but sometimes there will be a larger contact set. In either case, this contact set is called a *face* of the Grassmannian. In section 3.2 we will prove the following:

1.1 THE FACE THEOREM. *If all of the tangent planes to an oriented m-dimensional surface S in $\mathbb{R}^n$ lie in the same face of the Grassmannian $G(m, \mathbb{R}^n) \subset \mathbb{R}^N$, then S is area-minimizing.*

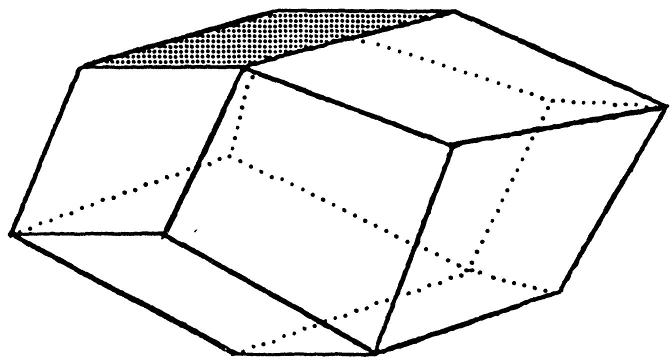


FIG. 1.1(1). A face of the Grassmannian (schematic).

Let's return for a moment to the Grassmannian $G(1, \mathbb{R}^2)$ of oriented lines through the origin in the plane. $G(1, \mathbb{R}^2)$ sits inside $\mathbb{R}^{\binom{2}{1}}$, another copy of $\mathbb{R}^2$. The coordinates of the line at angle θ to the x-axis are the oriented projections of a unit length from that line onto the axes, namely $(\cos\theta, \sin\theta)$. Hence $G(1, \mathbb{R}^2)$ is just the unit circle in $\mathbb{R}^2$. Its faces are all single points. The Face Theorem 1.1 just says that if all of the tangents to a curve are the same, i.e., if the curve is a straight line, then the curve is length-minimizing.

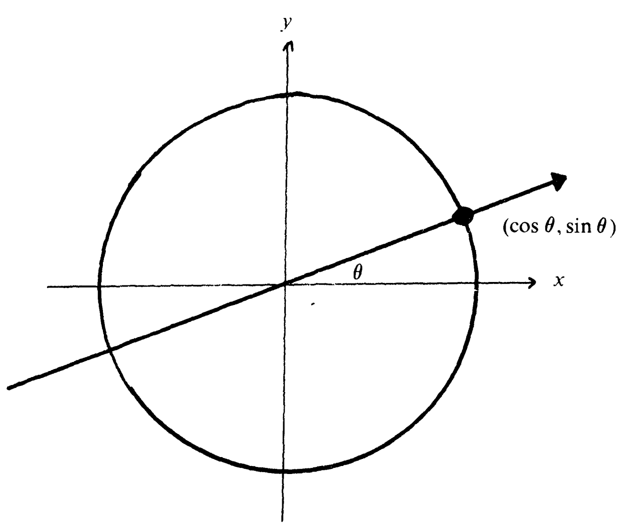


FIG. 1.1(2). The Grassmannian of lines in $\mathbb{R}^2$ is represented by the unit circle.

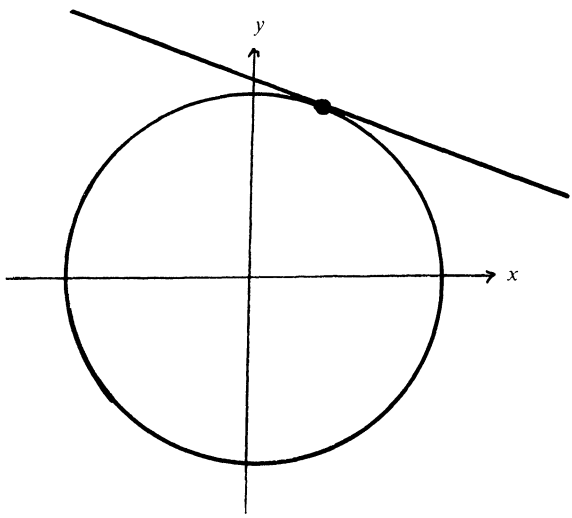


FIG. 1.1(3). Each face of $G(1, \mathbf{R}^2)$ is a single point.

Actually whenever $m = 1$ or $n - 1$, the Grassmannian $G(m, \mathbb{R}^n)$ is just the $(n-1)$-dimensional sphere in $\mathbb{R}^n$. The faces are all single points. The Face Theorem just says that straight lines and hyperplanes are area-minimizing.

In general dimensions, the Grassmannian $G(m, \mathbb{R}^n)$ is an $m(n-m)$-dimensional submanifold of the unit sphere in $\mathbb{R}^{\binom{n}{m}}$, and it has some larger faces. These faces give rise to interesting area-minimizing surfaces.

The main reference is the beautiful foundational paper of R. Harvey and H. B. Lawson [**12**]. There is also a nice survey by Harvey [**11**].

Before discussing the theory, we begin with some examples and applications.

2. Examples and applications.

2.1 *Complex geometry.* Let us identify $\mathbf{R}^{2n}$ with $\mathbf{C}^n$ and suppose $1 \leqslant m \leqslant n$. One of the faces of the Grassmannian $G(2m, \mathbf{R}^{2n})$ turns out to be the set of m-complex-dimensional complex planes in $\mathbf{C}^n$. Since the tangent planes to a complex analytic variety are all complex planes, it follows that a compact portion of a complex analytic variety is area-minimizing. For example, the complex curve

$$C = \left\{(z, w) \in \mathbf{C}^2 \colon w^2 = z^3, \qquad |z|^2 + |w|^2 \leqslant 1\right\}$$

is a two-real-dimensional, area-minimizing surface in the unit ball in $\mathbf{R}^4$, bounded by the 2-3 torus knot

$$\partial C = \left\{(z, w) \in \mathbf{C}^2 \colon z = e^{i\theta}, \qquad w = e^{(3/2)i\theta}, \qquad 0 \leqslant \theta \leqslant 4\pi\right\}.$$

C is a smooth submanifold with boundary, except for the isolated singularity at the origin. H. Federer's observation [**8**], that all complex analytic varieties are area-minimizing, provided some of the first examples of singularities in area-minimizing surfaces.

2.2 *Special Lagrangian geometry* [**12**, Chapter III]. Again identify $\mathbf{R}^{2n} \cong \mathbf{C}^n$. The *special Lagrangian* face of $G(n, \mathbf{R}^{2n})$ consists of the real plane $\xi_0 = \{z \in \mathbf{C}^n: \operatorname{Im} z = 0\}$ together with all of its images under the group SU_n of special unitary maps:

$$SU_n = \{n \times n \text{ complex matrices } A: A\bar{A}^t = I \text{ and } \det A = 1\}.$$

This special Lagrangian face is an $(n+2)(n-1)/2$-dimensional subset of the n^2-dimensional Grassmannian $G(n, \mathbb{R}^{2n})$. By comparison, for n even, the face of n-real-dimensional complex planes has dimension $n^2/2$.

An n-dimensional surface, with tangent planes in the special Lagrangian face, is called *special Lagrangian*. There are lots of these surfaces, and they must all be area-minimizing. For example, if M is any compact 2-dimensional minimal surface with boundary in R^3, then its normal bundle,

$$S = \{(x, \nu) \in \mathbf{R}^6 : x \in M \text{ and } \nu \text{ is a vector normal to } M \text{ at } x\},$$

turns out to be a 3-dimensional special Lagrangian surface in $\mathbf{R}^6$. Here M is not assumed to be area-minimizing, merely minimal. This term *minimal* means that the first variation of the area is 0, or equivalently, that the mean curvature is 0, or equivalently, that a parametrization in isothermal coordinates is harmonic. Even branch points in M are allowed, producing branch points in S.

B. Cheng [**3**] recently has discovered some interesting special Lagrangian surfaces. Let $m \geqslant 3$. By thinking of SU_m as a set of $m \times m$ complex matrices, we may view SU_m as a subset of the sphere of radius $\sqrt{m}$ in $\mathbf{C}^{m^2} \cong \mathbf{R}^{2m^2}$. Cheng has observed that the cone C over SU_m

$$C = \{tx: x \in SU_m, 0 \leqslant t \leqslant 1\}$$

is an m^2-dimensional special Lagrangian surface in $\mathbf{R}^{2m^2}$, bounded by SU_m.

For m odd, this cone C has an interesting feature: the singularity at the origin is not real analytic. Indeed, if $x \in C - \{0\}$, then $tx \in C$ if and only if $t \geqslant 0$; whereas, if a real-analytic variety contains a half line, it must contain the whole line. This is possibly the simplest known example of an area-minimizing surface which is not real analytic.

A function $f: \mathbf{R}^n \to \mathbf{R}^n$ whose graph is special Lagrangian satisfies a beautiful analog of the Cauchy-Riemann equations (see [**12**, III.2]).

2.3. *The angle conjecture* [**2**, Open problem $\sharp$ 5.8]. A basic question in the study of singularities in m-dimensional area-minimizing surfaces asks which pairs D, D' of oriented unit discs, centered at the origin, are area-minimizing. No pairs of 2-dimensional discs in $\mathbb{R}^3$ are area-minimizing, as Figure 2.3(1) indicates. It happens that certain pairs of 2-dimensional discs in $\mathbb{R}^4$ are area-minimizing. For the general case of m-dimensional discs in $\mathbb{R}^n$, the answer depends on the geometric relationship between the two discs, which can be described by m angles, $0 \leqslant \theta_1 \leqslant \theta_2 \leqslant \cdots \leqslant \theta_m$, as follows.

Choose unit vectors $v_1 \in D$, $v_1' \in D'$ to minimize the angle θ_1 between them. Choose unit vectors v_2 perpendicular to v_1 in D and v_2' perpendicular to v_1' in D', to minimize the angle θ_2 between them. Continue. Finally, choose v_m to complete an oriented orthonormal basis for D, v_m' to complete an oriented orthonormal basis for D', and let θ_m be the angle between them.

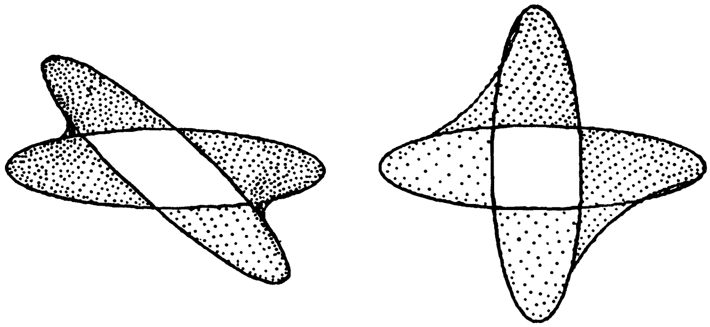


FIG. 2.3(1). Comparsion surfaces of less area than a pair of discs in $\mathbb{R}^3$.

The *angle conjecture* says that the pair of discs is area-minimizing if and only if the largest angle is less than or equal to the sum of the others:

$$\theta_m \leqslant \theta_1 + \cdots + \theta_{m-1}.$$

Note that for a pair of discs in $\mathbb{R}^3$, $m = 2$, the smaller angle θ_1 is always 0, and the condition $\theta_2 \leqslant \theta_1 = 0$ fails to hold.

D. Nance **[16]** recently has proved this condition sufficient by showing that such a pair of planes lie in a common face of the Grassmannian. The face turns out to be a clever generalization of the special Lagrangian face, based on the quaternions.

More recently G. Lawlor **[14]** has proved the condition necessary. He shows that planes violating the condition are not area-minimizing by providing comparison surfaces of less area. These comparison surfaces are special Lagrangian, and hence actually area-minimizing.

The proof of the angle conjecture is now complete.

2.4. *The classification problem.* What are all of the faces of the Grassmannian $G(m, \mathbf{R}^n)$? This classification problem has been solved only in low dimensions.

For $m = 1$ or $m = n - 1$, the only faces are single points, and the associated area-minimizing surfaces are just lines and hyperplanes.

For $m = 2$, every face consists of the complex lines in some even-dimensional subspace of $\mathbf{R}^n$ (under some orthogonal complex structure). For example, one face is a round $\mathbf{S}^2 \cong \mathbf{C}P^1$ of complex lines in some $\mathbf{R}^4 \cong \mathbf{C}^2$. For $m = n - 2$, a complementary classification holds. Thus for dimension or codimension 2, nothing but classical complex geometry appears.

The faces of the 9-dimensional Grassmannian $G(3, \mathbf{R}^6)$, the first nonclassical case, were classified in 1982 by J. Dadok and R. Harvey **[4]** and by **[15]**. In addition to the singleton and round $\mathbf{S}^2$ faces inherited from lower dimensions, two new types of faces appeared. First, there are large, five-dimensional, special Lagrangian faces as previously described, with lots of associated three-dimensional, area-minimizing surfaces in $\mathbf{R}^6$. Second, there are doubleton faces, consisting of certain pairs of three-planes. The pairs which occur are precisely the ones needed to prove the sufficiency of the Angle Conjecture condition for a pair of planes to be area-minimizing.

The last Grassmannian to be classified to date is the 12-dimensional Grassmannian $G(3, \mathbf{R}^7)$ **[13]**. In addition to five types of faces inherited from lower dimen-

sions, five new types of faces appear. The largest faces, eight-dimensional *associative* faces, produce lots of 3-dimensional area-minimizing surfaces in $\mathbf{R}^7$. They are called *associative* because of a relationship with the partial associativity of octonion multiplication.

A second new type of face consists of two round $\mathbb{S}^2$'s which intersect in a single point. The other three new types of faces are $\mathbb{S}^1$'s, $\mathbb{S}^2$'s, and $\mathbb{S}^3$'s which are not round. They provide the first examples of faces which are not totally geodesic in the Grassmannian.

3. Theory of calibrations

3.1. DEFINITIONS. Recall that a *face* of the Grassmannian $G(m, \mathbb{R}^n) \subset \mathbb{R}^{\binom{n}{m}}$ is defined as the initial contact set with a hyperplane translated in from infinity. The hyperplane is the level set of some linear function φ on $\mathbb{R}^{\binom{n}{m}}$.

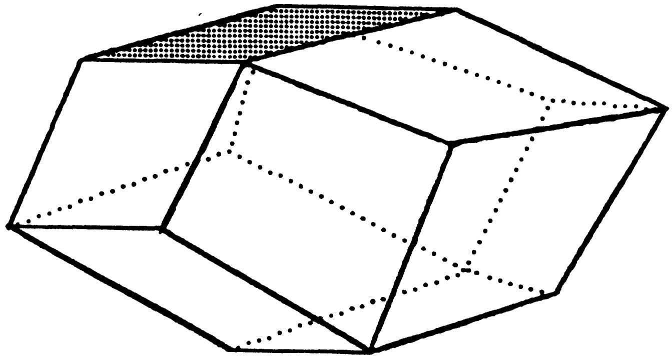


FIG. 3.1(1). A face is a level-set of a linear function φ.

It is easy to check that the linear functions on $\mathbb{R}^{\binom{n}{m}}$ are precisely the constant-coefficient, differential m-forms on $\mathbb{R}^n$. Indeed, an orthonormal basis for $\mathbb{R}^{\binom{n}{m}}$ is given by the axis m-planes in $G(m, \mathbb{R}^n) \subset \mathbb{R}^{\binom{n}{m}}$:

$$\left\{ e_{i_1} \wedge \cdots \wedge e_{i_m} : 1 \leqslant i_1 < \cdots < i_m \leqslant n \right\};$$

and the dual basis for the constant-coefficient differential m-forms is given by

$$\left\{ dx_{i_1} \wedge \cdots \wedge dx_{i_m} : 1 \leqslant i_1 < \cdots < i_m \leqslant n \right\}.$$

Thus a face of the Grassmannian $G(m, \mathbb{R}^n) \subset \mathbb{R}^{\binom{n}{m}}$ is the set of points in $G(m, \mathbb{R}^n)$ where an m-form φ attains its maximum value

$$\|\varphi\|^* = \max\{\varphi(\xi) : \xi \in G(m, \mathbf{R}^n)\},$$

called the *comass* of φ. A (constant-coefficient)m-form φ, normalized to have comass 1, is called a *calibration*. The faces of the Grassmannian are precisely the sets of maxima of calibrations φ.

3.2. THE FACE THEOREM [**12**, Introduction]. *Let S be an m-dimensional surface with boundary in $\mathbf{R}^n$ such that every tangent plane to S lies in a common face of the Grassmannian. Then S is area-minimizing, i.e., no other surface T with the same boundary has less area.*

Proof. Since there is a face of the Grassmannian containing every tangent plane to S, there is a calibration φ such that $\varphi(\xi) \leqslant 1$, with equality for ξ tangent to S. Hence $\int_S \varphi =$ area S, while $\int_T \varphi \leqslant$ area T. Since φ is a constant-coefficient differential form, $d\varphi = 0$. Hence it follows by Stokes's Theorem that $\int_S \varphi = \int_T \varphi$, because S and T have the same boundary. Therefore,

$$\text{area } S = \int_S \varphi = \int_T \varphi \leqslant \text{area } T,$$

as claimed.

3.3. *Examples of calibrations.* This section gives explicitly the calibrations that yield the faces of Grassmannians heralded at the beginning of the article.

Complex lines in $\mathbf{C}^n$ comprise the maxima of the Kähler form

$$\begin{aligned} \omega &= \frac{i}{2}\left(dz_1 \wedge \overline{dz_1} + \cdots + dz_n \wedge \overline{dz_n}\right). \\ &= dx_1 \wedge dy_1 + \cdots + dx_n \wedge dy_n. \end{aligned}$$

Similarly, complex m-planes in $\mathbf{C}^n$ comprise the maxima of $\omega^m/m!$ (here normalized to have maximum value 1). These facts follow immediately from Wirtinger's Inequality (cf. [**6**, 1.8.2, 5.4.19]).

Special Lagrangian m-planes in $\mathbf{R}^{2m} \cong \mathbf{C}^m$ comprise the maxima of the *special Lagrangian form*

$$\begin{aligned} \varphi &= \operatorname{Re} dz = \operatorname{Re} dz_1 \wedge \cdots \wedge dz_n. \\ &= \operatorname{Re}(dx_1 + i\,dy_1) \wedge \cdots \wedge (dx_n + i\,dy_n). \end{aligned}$$

For example, in $\mathbf{R}^6$,

$$\varphi = dx_1 \wedge dx_2 \wedge dx_3 - dx_1 \wedge dy_2 \wedge dy_3 - dy_1 \wedge dx_2 \wedge dy_3 - dy_1 \wedge dy_2 \wedge dx_3.$$

R. Harvey and H. B. Lawson [**12**] discovered this and other beautiful calibrations in 1976.

The *Nance forms*, used to prove half the Angle conjecture, generalize the special Lagrangian form by replacing i by more general imaginary unit quaternions $u_1, \ldots, u_m$:

$$\varphi = \operatorname{Re}(dx_1 + u_1 dy_1) \wedge \cdots \wedge (dx_m + u_m dy_m).$$

Given a pair of m-planes whose characterizing angles satisfy the criterion

$$\theta_m \leqslant \theta_1 + \cdots + \theta_{m-1},$$

a calibration can be constructed as follows. The inequality lets us construct a polygon in the 2-sphere of imaginary unit quaternions with sides $\theta_1, \ldots, \theta_m$, as shown in FIGURE 3.3(1). Now the $u_1, \ldots, u_m$ for defining φ are just the exterior poles of the corresponding arcs. (For example, if θ_4 lies on the equator, and the polygon lies above the equator, u_4 is the south pole.) The whole proof [**16**] is short and elegant.

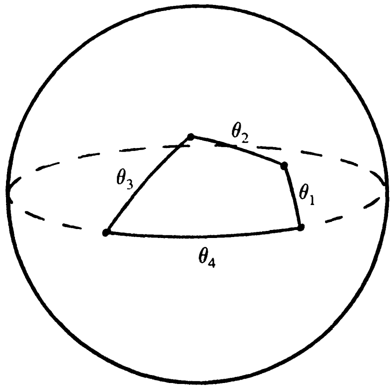


FIG. 3.3(1). Admissible angles form the side-lengths of a spherical polygon.

J. Dadok and R. Harvey [**5**] recently have used spinors to construct new calibrations.

3.4. *Calibrations on manifolds.* The following theorem shows how the theory of calibrations, presented above in Euclidean space, extends to general Riemannian manifolds.

THEOREM [**12**, II.4]. *Let S be an oriented m-dimensional surface, with or without boundary, in a smooth Riemannian manifold M. Let φ be a smooth closed differential m-form on M. Suppose that for all $x \in M$, for all unit m-planes ξ, $\varphi(x)(\xi) \leqslant 1$, with equality if ξ is tangent to S at x. Then no oriented surface T homologous to S has less area.*

The proof is the same as for Theorem 3.2.

3.5. *Least-area representatives of homology classes.* Theorem 3.4 sometimes can be applied to find explicitly surfaces of least area in homology classes in Riemannian manifolds. If the manifold M is complex and the Kähler form ω is closed, i.e., if M is a Kähler manifold, then by Wirtinger's inequality the normalized powers $\omega^p/p!$ of ω calibrate all complex subvarieties. Hence these complex subvarieties minimize area in their homology classes. For example, every complex algebraic subvariety of complex projective space $\mathbf{C}P^n$ is homologically area-minimizing.

In 1972 M. Berger [**1**, §6] established an analog of Wirtinger's inequality for a quaternionic form Ω. Consequently, if M is "quaternionic-Kähler," so that M admits a closed quaternionic form Ω, then the normalized powers of Ω calibrate all quaternionic subvarieties. Hence these quaternionic subvarieties minimize area in their homology classes. For example, every quaternionic subvariety of quaternionic projective space is homologically area-minimizing.

A. T. Fomenko began the search for homologically area-minimizing surfaces in compact Lie groups and symmetric spaces. He proved for example that inside the Lie group $M = SU_n$, the subgroup SU_2 is homologically area-minimizing [**9**, Corollary 8]. Dao Chong Thi [**20**] used methods equivalent to calibrations (cf. [**12**, V.1]) to obtain this and other new examples. It is still an open question whether SU_m is homologically area-minimizing in SU_n for $2 < m < n$.

As a final example, let M be the Grassmannian $G(k, \mathbb{R}^p)$ of oriented k-planes through the origin in $\mathbb{R}^p$. Sub-Grassmannians, as totally geodesic submanifolds, are excellent candidates for least-area representatives of homology classes. For $k = 2$, $G(2, \mathbb{R}^p)$ is a Kähler manifold, and the various sub-Grassmannians $G(2, \mathbb{R}^q)$ $(q < p)$ are complex submanifolds and hence homologically area-minimizing. For $k = 4$, $G(4, \mathbb{R}^p)$ is a quaternionic-Kähler manifold, and the various sub-Grassmannians $G(4, \mathbb{R}^p)$ $(q < p)$ are quaternionic submanifolds and, hence, homologically area-minimizing. More recently, by using new calibrations related to the Euler form, H. Gluck, W. Ziller, and the author **[10]** have proven that for any even k, $G(k, \mathbb{R}^q)$ is homologically area-minimizing in $G(k, \mathbb{R}^p)$.

However, $G(2, \mathbb{R}^4)$ is not homologically area-minimizing in $G(3, \mathbb{R}^6)$, despite the fact that it is totally geodesic and homologically nontrivial. (The inclusion of $G(2, \mathbb{R}^4)$ in $G(3, \mathbb{R}^6)$ is given by $\zeta \to \zeta \wedge e_5$.)

The analysis leads to an open question involving the special orthogonal group

$$SO_3 = \{3 \times 3 \text{ real matrices } A\colon AA^t = I \text{ and } \det A = 1\}.$$

Question: Does the cone C over SO_3,

$$C = \{tx \in \mathbb{R}^9 : x \in SO_3, 0 \leqslant t \leqslant 1\},$$

minimize area among all oriented four-dimensional surfaces bounded by SO_3?

This stubborn question seems to require a new idea for its solution.

Added in proof. G. Lawlor has proved that the cone C over SO_3 is area-minimizing by a powerful new method. However, many questions remain open.

The author acknowledges partial support from the National Science Foundation.

REFERENCES

1. M. Berger, Du côté de chez Pu, *Ann. scient. Éc. Norm. Sup.*, (4)5 (1972) 1–44.
2. John E. Brothers, ed., Some open problems in geometric measure theory..., *Geometric Measure Theory and the Calculus of Variations* (William K. Allard and Frederick J. Almgren, Jr., ed.), Proc. Symp. Pure Math., 44 (1986) 441–464.
3. B. Cheng, Ph.D. Thesis, MIT, 1987
4. Jiri Dadok and Reese Harvey, Calibrations on $\mathbf{R}^6$, *Duke Math. J.*, 50(1983) 1231–1243.
5. ______, in preparation.
6. Herbert Federer, *Geometric Measure Theory*, Springer-Verlag, New York, 1969.
7. ______, Real flat chains, cochains, and variational problems, *Ind. U. Math. J.*, 24(1974) 351–407.
8. ______, Some theorems on integral currents, *Trans. Am. Math. Soc.*, 117(1965) 43–67.
9. A. T. Fomenko, Minimal compacta in Riemannian manifolds and Riefenberg's conjecture, *Izv. Akad. Nauk SSSR Serv. Mat.*, 36(1972); English translation, *Math. USSR Izv.*, 6(1972) 1037–1066.
10. Herman Gluck, Frank Morgan, and Wolfgang Ziller, Calibrated geometries in Grassmann manifolds, preprint.
11. Reese Harvey, Calibrated geometries, *Proc. Int. Cong. Math.*, 1983.
12. Reese Harvey and H. Blaine Lawson, Jr., Calibrated geometries, *Acta Math.*, 148(1982) 47–157.
13. Reese Harvey and Frank Morgan, The faces of the Grassmannian of three-planes in $\mathbf{R}^7$, *Invent. Math.*, 83(1986) 191–228.
14. Gary Lawlor, The angle criterion, preprint.
15. Frank Morgan, The exterior algebra $\wedge^k \mathbf{R}^n$ and area minimization, *Lin. Alg. App.*, 66(1985) 1–28.
16. Dana Nance, Sufficient conditions for a pair of n-planes to be area-minimizing, *Math. Ann.*, 279(1987) 161–4.
17. Yoshihiro Ohnita and Hiroyuki Tasaki, Uniqueness of certain 3-dimensional homologically volume minimizing submanifolds in compact simple Lie groups, *Tsukuba J. Math.*, 10(1986) 11–16.

18. Hiroyuki Tasaki, Certain minimal or homologically volume minimizing submanifolds in compact symmetric spaces, *Tsukuba J. Math.*, 9(1985) 117–131.
19. ______, Quaternionic submanifolds in quaternionic symmetric spaces, *Tôhoku Math. J.*, 38(1986) 513–38.
20. Dao Chong Thi, Minimal real currents on compact Riemannian manifolds, *Izv. Akad. Nauk. SSSR Ser. Mat.*, 41(1977) No. 4, 853–867.
21. Hiroyuki Tasaki, Calibrated geometries in quaternionic Grassmannians, preprint.

Constructing Isospectral Manifolds

ROBERT BROOKS, *University of Southern California**

ROBERT BROOKS is an Associate Professor of Mathematics at the University of Southern California. He received his Ph.D. from Raoul Bott at Harvard, and was awarded an Alfred P. Sloan fellowship in 1984. His research interests include differential geometry and topology and the geometry of the Laplace operator.

Let M be a compact Riemannian manifold. The Laplace-Beltrami operator (or Laplacian, for short) is the operator

$$\Delta(f) = -\operatorname{div}(\operatorname{grad} f).$$

One way of understanding this operator is by the Heat Equation. If $f_t(x)$ denotes the temperature at time t and a point x, and if we assume that heat travels in the coolest direction, then $f_t(x)$ satisfies the equation

$$\frac{\partial}{\partial t}f + \Delta f = 0$$

as one sees from the usual physical interpretation of div and grad.

Now suppose that $f_0(x)$ is an eigenfunction of Δ with eigenvalue λ, that is, f_0 satisfies

$$\Delta(f_0) = \lambda \cdot f_0.$$

Then the function $f_t(x) = e^{-\lambda t}f_0(x)$ will satisfy the Heat Equation. One may think of f_t describing a "heat wave" with "frequency" $e^{-\lambda t}$.

Standard results in partial differential equations tell one that there are countably many eigenvalues λ_i and for each λ_i a finite-dimensional family of eigenfunctions f_i, so that λ_i and f_i satisfy

$$\Delta(f_i) = \lambda_i f_i.$$

Furthermore, the λ_i's are positive, and tend to ∞ as $i \to \infty$. The collection of values $\{\lambda_i\}$, together with the multiplicities with which they occur, is called the spectrum of M.

An arbitrary function $f(x)$ may then be decomposed into its "fundamental heat waves"

$$f(x) = \sum a_i f_i(x), \qquad \Delta(f_i) = \lambda_i(f_i)$$

in exactly the same way that the motion of a string decomposes into its fundamental overtones, the sine and cosine functions of Fourier analysis.

If one considers that sound satisfies a similar equation to that of heat, then one may regard the λ_i's as the sounds emitted by M, if one struck M with a mallet.

---

*Partially supported by NSF grant DMS-8501300

A natural question, in the best tradition of Riemannian geometry, is to ask for the geometric significance of the numbers λ_i, whose existence we found purely from analysis. It is hard to imagine this question not being asked even by the more speculative-minded of the Pythagoreans, but in the 20th century this question took on increasing interest, as the tools to answer it were developed.

In the first half of this century, there was a great deal of speculation that the eigenvalues $\{\lambda_i\}$ might completely determine the geometry of M. This was perhaps fueled by the striking result of Hermann Weyl **[22]** that the eigenvalues determine such geometric invariants as the volume of M. In his beautiful survey article **[15]**, Mark Kac framed this train of thought in the question, "Can you hear the shape of a drum?"

In recent years, attention has been focused on examples which show this speculation to be somewhat overambitious. Let us say that two manifolds are *isospectral* if the eigenvalues of the Laplacian are the same for both M_1 and M_2. Then we will see below how one may give rather explicit constructions of isospectral manifolds which are not isometric.

The first example of isospectral manifolds which were not isometric was found by John Milnor in **[17]**. We will discuss his example briefly in § 1 below, but for the present let us say that his method was to reduce the isospectral problem for certain classes of manifolds to a number-theoretic problem that had a well-known solution. His method was taken up, after a long period, by Ikeda **[14]** and Vignéras **[21]**, who were able to find examples of isospectral manifolds in other classes, by similarly looking for solutions to a companion problem among classical examples.

It is fair to say that these examples all had the flavor of happy, if somewhat exotic, coincidences, and it seemed as if the search for isospectral manifolds was leading mathematicians into increasingly esoteric subjects in order to find such coincidences.

The situation changed dramatically in **[20]**. Here, Sunada showed how the phenomenon of isospectral manifolds could be understood in a systematic way. His insight was to reduce the problem to a straightforward problem in the theory of finite groups. This reduction is quite simple, and we will present a complete argument in § 2 below. Furthermore, the problem in group theory admitted large numbers of relatively straightforward solutions and was understood in a fairly systematic way. Indeed, this same group-theoretic problem arose in a basic problem in number theory—see **[24]** and **[25]**.

At this point, the construction of isospectral manifolds became fairly simple, and actually rather fun.

In this article, we would like to share these developments with you. We hope to share some of the excitement that occurs when certain kinds of objects which were thought not to exist—or at least to be extremely rare—can be constructed at home using scissors, paper, and tape.

The literature on the spectrum of the Laplacian is quite extensive, and we will not make an attempt to give a complete overview of it. This topic has been the subject of three previous review articles in the *American Mathematical Monthly* (**[9]**, **[15]**, **[23]**). A comprehensive bibliography concerning the spectrum of the Laplacian as of 1982 has been compiled by Bérard and Berger in **[1]**. Furthermore, the foundational material of the subject has been collected in at least two monographs, **[2]** and **[6]**, which we heartily recommend.

We would like to thank K. Layman and R. Perlis for their critical readings of this paper.

1. The Jacobi Inversion Formula

Among the simplest manifolds are the flat tori. To describe them, let us pick two points (x_1, y_1) and (x_2, y_2) in $\mathbb{R}^2$ which are linearly independent. Let Λ be the lattice generated by (x_1, y_1) and (x_2, y_2)—that is, all linear combinations $m \cdot (x_1, y_1) + n \cdot (x_2, y_2)$. Then the torus T^2 is determined by identifying points which differ by an element of Λ:

$$T^2 = \mathbb{R}^2/\Lambda.$$

Topologically, T^2 is the surface of a doughnut, as one may see by cutting out the parallelogram spanned by (x_1, y_1) and (x_2, y_2), and gluing opposite sides together. Furthermore, two lattices Λ_1 and Λ_2 give rise to the same torus if and only if Λ_2 differs from Λ_1 by a rotation. There is a 3-dimensional family of distinct flat tori.

In this section, we will prove:

THEOREM. *Any two flat tori which are isospectral are isometric.*

This theorem is basically the Jacobi inversion formula, which we will describe below. Although this theorem seems a little outside the mainstream of our discussion, because it is a theorem showing that certain manifolds *cannot* be isospectral, all of the ingredients of the subsequent theory can be found here, at least in some form.

A function $H_t(x, y)$ will be called a *fundamental solution of the Heat Equation* (or a *heat kernel*) if

(a) $$\left(\frac{\partial}{\partial t} + \Delta_x\right) H = 0$$

(b) $$\lim_{t \to 0} \int_M H_t(x, y) f(y)\, dy = f(x)$$

for any smooth function f on M.

From (a) and (b), we see that the general solution $f_t(x)$ of the heat equation with initial condition $f_0(x) = f$ is given by

$$f_t(x) = \int_M H_t(x, y) f(y)\, dy.$$

We will need the following properties of H_t:

(i) $H_t(x, y)$ is uniquely determined by (a) and (b).

(ii) If M is compact, and $\{f_i\}$ is an orthonormal basis of eigenfunctions with eigenvalues λ_i, then

$$H_t(x, y) = \sum_i e^{-\lambda_i t} f_i(x) f_i(y).$$

We now want to use (ii) to describe the λ_i's, but we need to eliminate the f_i's from the picture. To do this, let

$$\operatorname{tr}(H_t) = \int_M H_t(x, x)\, dx.$$

Then

$$\operatorname{tr}(H_t) = \int_M \sum e^{-\lambda_i t} f_i^2(x)\, dx = \sum e^{-\lambda_i t} \int_M f_i^2\, dx$$
$$= \sum e^{-\lambda_i t} \quad \text{(since the } f_i\text{'s are orthonormal)}.$$

Now we claim: $\operatorname{tr}(H_t)$ determines all of the λ_i's.

Proof. Assume $\lambda_0, \lambda_1, \ldots, \lambda_k$ have been found. Then λ_{k+1} is the largest value λ such that

$$\lim_{t \to \infty} \frac{\operatorname{tr}(H_t) - \sum_{i=0}^{k} e^{-\lambda_i t}}{e^{-\lambda t}} \quad \text{is finite.}$$

Now suppose that M' is a covering of M—this means that there is a map $\varphi\colon M' \to M$ which near every point is an isometry. Such a manifold M' is determined by a subgroup $\pi_1(M')$ of the fundamental group $\pi_1(M)$. When $\pi_1(M')$ is a normal subgroup of $\pi_1(M)$, M' is said to be a normal covering of M, and the quotient group $G = \pi_1(M)/\pi_1(M')$ acts on M' as a group of isometries of M'. G is called the covering group.

In this situation, we would like to solve for the heat kernel $H_t^M(x, y)$ of M in terms of $H_t^{M'}(x, y)$:

LEMMA. *$H_t^M(x, y) = \sum_{g \in G} H_t^{M'}(\tilde{x}, g \cdot \tilde{y})$ where $\tilde{x}$ and $\tilde{y}$ are points of M' covering x and y.*

Proof. It is evident that the right-hand side satisfies (a) and (b) above. The lemma now follows from the uniqueness statement of (i).

We now specialize this discussion to the case M a flat torus, that is

$$M = \mathbb{R}^2/\Lambda, \quad \Lambda \text{ a lattice}$$

and $M' = \mathbb{R}^2$. We observe that $H_t^{\mathbb{R}^2}(x, y)$ is given by the "Gaussian distribution"

$$H_t^{\mathbb{R}^2}(x, y) = \left(\frac{1}{4\pi t}\right) e^{|x-y|^2/4t}$$

so from the lemma, we see

$$H_t^M(x, x) = \left(\frac{1}{4\pi t}\right) \sum_{l \in \Lambda} e^{-|l|^2/4t}$$

and so

$$\operatorname{tr}\left(H_t^M\right) = \frac{\operatorname{area}(M)}{4\pi t}\left(\sum_{l \in \Lambda} e^{-|l|^2/4t}\right).$$

We may now argue as in the claim, but now letting $t \to 0+$ instead of $\to \infty$, to show that $\operatorname{tr}(H_t^M)$ determines area (M) and

$$\{|l| \colon l \in \Lambda\}.$$

Since (H_t^M) depends only on the eigenvalues $\{\lambda_i\}$, we conclude:

If two flat tori $M_1 = \mathbb{R}^2/\Lambda_1$ and $M_2 = \mathbb{R}^2/\Lambda_2$ are isospectral, then area $(M_1) = \text{area}(M_2)$, and $\{|l|: l \in \Lambda_1\} = \{|l|: l \in \Lambda_2\}$.

To complete the proof of the theorem we observe that a lattice in $\mathbb{R}^2$ is determined up to isometry by its shortest length, the shortest length of a linearly independent lattice point, and the area of the parallelogram they span. Since all of these are determined from $\mathrm{tr}(H_t^M)$, which in turn is determined from the λ_i's, we see that M is determined from the λ_i's, proving the theorem.

To complete our discussion, we observe that for a flat torus, we may explicitly solve for the λ_i's as follows: given a lattice Λ in $\mathbb{R}^2$, the dual lattice Λ^* is the set $l^* \in \mathbb{R}^2$ such that $l^* \cdot l$ is an integer for all $l \in \Lambda$. Then the eigenfunctions f_i are given by

$$f_i(x) = e^{2\pi i(l^* \cdot x)} \quad \text{for } l^* \in \Lambda^*$$

with eigenvalue $4\pi^2|l^*|^2$, so that

$$\mathrm{tr}\left(H_t^M\right) = \sum_{l^* \in \Lambda^*} e^{-4\pi^2|l^*|^2 t}$$

and we conclude the Jacobi inversion formula:

$$\sum_{l^* \in \Lambda^*} e^{-4\pi^2|l^*|^2 t} = \frac{\text{area}(M)}{4\pi t}\left(\sum_{l \in \Lambda} e^{-|l|^2/4t}\right).$$

It is evident that this argument generalizes to any dimension n. There were only two ingredients of this argument which were two-dimensional. The first of these was the expression for the heat kernel for $\mathbb{R}^2$. But the heat kernel for $\mathbb{R}^n$ differs only slightly from the heat kernel for $\mathbb{R}^2$. The second is the assertion that a lattice in $\mathbb{R}^2$ is determined by its area and lengths. This argument breaks down in higher dimensions, but we are left with the conclusion: if $T_1 = \mathbb{R}^n/\Lambda_1$, and $T_2 = \mathbb{R}^n/\Lambda_2$ are isospectral, then the volumes of T_1 and T_2 are the same, and the set of lengths of lattice points are the same.

Milnor observed in [7] that there were two lattices in $\mathbb{R}^{16}$ which were well known to number theorists to have this property, and he concluded that there were two 16-dimensional flat tori which are isospectral.

One can generalize this argument for any manifold for which you can write down the heat kernel explicitly—for instance, symmetric spaces of noncompact type. For the hyperbolic plane, the same argument gives (after some more elaborate calculation) the Selberg trace formula ([**16**] and [**18**]), which says that, when S is a surface of constant curvature -1, the eigenvalues of S determine the lengths of all the closed geodesics of S, as well as the area of S.

2. The Sunada Trace Formula

In this section, we will show how to extend the ideas of the last section in order to construct isospectral manifolds. In fact, the striking thing about this argument is that it is in fact a simplification of the previous argument, since it will involve only finite groups, and hence finite sums. Furthermore, it will not involve computing any complicated integrals or functions.

Let us start with a normal covering M' of M, which we now assume to be a finite covering. As before, we wish to solve for $H_t^M(x, y)$ in terms of $H_t^{M'}(x, y)$, and as in the lemma, we have

$$H_t^M(x, y) = \sum_{g \in G} H_t^{M'}(\tilde{x}, g\tilde{y}).$$

Setting $x = y$ and integrating over M, we have

$$\operatorname{tr}\left(H_t^M\right) = \int_M H_t^M(x, x)\, dx = \frac{1}{\#(G)} \sum_{g \in G} \int_{M'} H_t^{M'}(\tilde{x}, g\tilde{x})\, d\tilde{x}.$$

We now observe that the terms in the right-hand side are unchanged when g is replaced by a conjugate hgh^{-1} of g, since:
(a) for any isometry h of M',

$$H_t^{M'}(h\tilde{x}, h\tilde{y}) = H_t^{M'}(\tilde{x}, \tilde{y})$$

(this follows from the uniqueness statement (i) of § 2) so that

(b) $$\int_{M'} H_t^{M'}(\tilde{x}, hgh^{-1}\tilde{x})\, d\tilde{x} = \int_{M'} H_t^{M'}(h^{-1}\tilde{x}, gh^{-1}\tilde{x})\, d\tilde{x}$$
(by (a))

$$= \int_{M'} H_t^{M'}(\tilde{x}, g(\tilde{x}))\, d\tilde{x} \quad \text{(changing variables)}.$$

We conclude that

$$\operatorname{tr}\left(H_t^M\right) = \sum_{[g]} \frac{\#([g])}{\#(G)} \int_{M'} H_t^{M'}(\tilde{x}, g\tilde{x})\, d\tilde{x},$$

where $[g]$ denotes the conjugacy class of g in G.

Now suppose we have a tower of coverings

$$\begin{tikzcd} M' \arrow[d, "H"'] \arrow[dd, bend left, "G"] \\ M_1 \arrow[d] \\ M_0 \end{tikzcd}$$

where M' is a normal covering of M_1 (with covering group H), and also of M_0 (with covering group G), but M_1 is not necessarily a normal covering of M_0.

Then we have from the discussion above that

$$\operatorname{tr}\left(H_t^{M_1}\right) = \sum_{[h] \in H} \frac{\#([h])}{\#(H)} \int_{M'} H_t(\tilde{x}, h\tilde{x})\, d\tilde{x},$$

but we may now use the symmetries of G to observe that if h and h' are conjugate in G, then

$$\int_{M'} H_t(\tilde{x}, h\tilde{x})\, d\tilde{x} = \int_{M'} H_t(\tilde{x}, h'\tilde{x})\, d\tilde{x}$$

so that

$$\operatorname{tr}\left(H_t^{M_1}\right) = \sum_{[g]\in G} \frac{\#([g]\cap H)}{\#(H)} \int_{M'} H_t(\tilde{x}, g\tilde{x})\, d\tilde{x}.$$

Now the point is that the integrals on the right-hand side, as uncomputable as they may be, depend only on G, and not on H. Only the coefficients depend on H. So if we can find two subgroups H_1 and H_2 of G satisfying

$$\text{for all } g\in G,\ \#([g]\cap H_1) = \#([g]\cap H_2) \tag{†}$$

then the corresponding manifolds M_1 and M_2 will have identical $\operatorname{tr}(H_t)$ and hence will be isospectral.

We have thus proved:

THEOREM ([**20**]). *Let us choose groups G, H_1 and H_2 satisfying* (†), *and M_0, M_1, M_2, M' a diagram of covering space as follows*:

$$\begin{tikzcd} & M' \arrow[dl, "H_1"'] \arrow[dd, "G"] \arrow[dr, "H_2"] & \\ M_1 \arrow[dr] & & M_2 \arrow[dl] \\ & M_0 & \end{tikzcd}$$

Then M_1 is isospectral to M_2.

It remains to check that M_1 is not isometric to M_2. If H_1 is conjugate to H_2 in G (or even in a larger group of isometries of M') then of course M_1 will be isometric to M_2. But it is easy to check that if H_1 is not conjugate to H_2, and the metric on M_0 is sufficiently "bumpy" so that M' has no extra isometries not in G, then M_1 is not isometric to M_2.

3. Some Finite Groups

It now remains to show that there are examples of groups (G, H_1, H_2) satisfying the condition (†). Such a condition may seem strange to an analyst, but such groups are known in abundance to finite group theorists. We present below an arsenal of examples, taken from [**3**], [**12**], and [**20**]:

Example 1. $Z/8\times(Z/8)^*$.

We consider ordered pairs (a, b) with $b\in Z/8$, $a = 1, 3, 5$, or 7, with multiplication rule $(a, b)\cdot(a', b') = (aa', a'b + b')$.

The group G thus defined is the semi-direct product of $Z/8$ and $(Z/8)^*$, and has 32 elements. Let $H_1 = \{(1,0),(3,0),(5,0),(7,0)\}$ and $H_2 = \{(1,0),(3,4),(5,4),(7,0)\}$. It is easy to check that (G, H_1, H_2) satisfies (†), and that H_1 is not conjugate to H_2 in G.

Example 2. $G = SL(n, Z/p)$, $n\geqslant 3$,

$$H_1 = \begin{pmatrix} * & * & * & * \\ 0 & & & \\ 0 & & X & \\ 0 & & & \end{pmatrix} \qquad H_2 = \begin{pmatrix} * & 0 & 0 & 0 \\ * & & & \\ * & & X & \\ * & & & \end{pmatrix}.$$

To see that $\#([g] \cap H_1) = \#([g] \cap H_2)$, notice that g will appear in H_1 if and only if

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

is an eigenvector of H_1, and hence the number of times $[g]$ meets H_1 depends only on the number of eigenvectors g has. Since this is the same for g and its transpose, $\#([g] \cap H_1) = \#([g] \cap H_2)$. In this example, note that H_1 is conjugate to H_2 under the outer automorphism $g \to (g^t)^{-1}$, but that if $n \geqslant 3$, H_1 is not conjugate to H_2 in G. To see this last, note that if $n \geqslant 3$, the elements of H_2 do not have a common eigenvector (when $n = 2$, the vector $\binom{0}{1}$ is an eigenvector of all of H_2).

Example 3. Let G be the symmetric group on p^3 letters, and let $H_1 = (Z/p) \oplus (Z/p) \oplus (Z/p)$ and H_2 be the "mod p Heisenberg group"

$$H_2 = \left\{ \begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix} : a, b, c, \in Z/p \right\}.$$

We may view H_1 and H_2 as subgroups of G by the Cayley-Sylvester theorem, which says that any finite group H is a subgroup of the symmetric group on $\#(H)$ letters. We claim that both H_1 and H_2 share the following property: any nonzero element of H_i, $i = 1, 2$, has order exactly p. For H_1 this is obvious, and for H_2 this can be verified from the formula

$$\begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix}^k = \begin{pmatrix} 1 & k \cdot a & k \cdot b + \dfrac{(k)(k-1)}{2} ac \\ 0 & 1 & k \cdot c \\ 0 & 0 & 1 \end{pmatrix}.$$

We now claim that $\#([g] \cap H_1) = \#([g] \cap H_2)$ for all $g \in G$. But the conjugacy class of g is determined by its decomposition into cycles, and for $[g]$ to meet either H_1 or H_2, it must consist of p^2 p-cycles, in which case $\#([g] \cap H_1) = \#([g] \cap H_2) = p^3 - 1$, or $g = id$, in which case $\#([g] \cap H_1) = \#([g] \cap H_2) = 1$. Note that not only is H_1 not conjugate to H_2, but H_1 is not even abstractly isomorphic to H_2, since H_1 is abelian and H_2 is not.

Example 4. We may extend example 3 by letting H_1 and H_2 be any two groups of order p^a, such that each nonzero element has order exactly p, and letting G be the symmetric group of order p^a. For instance, taking a large and considering the H_i's to be direct sums of Heisenberg groups and $(Z/p)^3$'s, we see that there are G's with arbitrarily many subgroups H_i, all pairwise satisfying (†) and all pairwise distinct.

The above collection of examples will convince the reader that groups satisfying (†) exist in abundance. Somewhat surprising, then, is the fact that groups satisfying (†) can be more or less systematically understood. See **[12]** for a discussion.

4. Isospectral Manifolds

The theorem of § 2 and the examples of § 3 have now amply illustrated:

THEOREM. *There exist manifolds M_1 and M_2 which are isospectral but not isometric.*

Proof. It suffices to take a triple of groups (G, H_1, H_2) satisfying (†) and a sufficiently bumpy manifold M_0 with a surjective homomorphism φ: $\pi_1(M_0) \to G$. We may then take M_i, $i = 1, 2$ to be the covering of M_0 with fundamental group $\varphi^{-1}(H_i)$.

For instance, we may use the fact that every finitely presented group occurs as the fundamental group of a compact manifold. Then, using Example 3, we may find M_1 and M_2 which are not even homeomorphic—they even have different fundamental groups.

An important class of manifolds to which the Theorem of § 2 does not apply directly is the class of Riemann surfaces—surfaces equipped with metrics of constant curvature. The constant curvature condition restricts how "bumpy" one can make the metric on M_0, so it is not clear that M_1 and M_2 so constructed can be made non-isometric. Fortunately, surfaces are sufficiently simple to allow one to see directly that they are not isometric. It is an interesting challenge to find examples which are particularly simple—that is, with smallest possible genus.

The best we can do along this direction is:

THEOREM ([4], [5]).
(a) *For each genus $g \geqslant 4$, there are Riemann surfaces S_1 and S_2 of genus g which are isospectral but not isometric.*
(b) *There exist surfaces S_1 and S_2 of variable negative curvature and genus* 3 *which are isospectral but not isometric.*

The case of genus 5 and genus $\geqslant 7$ was done by Buser in [5]. Here we will sketch the construction of genus 4.

In order to construct examples of low genus, one must first make a nice choice of groups (G, H_1, H_2). Since the area of a Riemann surface is $4\pi(g-1)$, by the Gauss-Bonnet theorem, and since areas multiply under coverings, we want to choose the genus of M_0 small and the index of H_1 in G small.

In fact, we will want to choose the genus of M_0 so small that M_0 must be singular. We will then have the further problem of guaranteeing that the singularities of M_0 smooth out in the coverings M_1 and M_2.

For our construction, we will take the group $G = SL(3, Z/2)$. We choose this because in this case $[G : H_1] = [G : H_2] = 7$, and this can be shown to be the smallest index which can occur for groups satisfying (†).

For our base manifold M_0 we will choose a torus with one singular point of order k, where k will be determined in a moment. Then the fundamental group (suitably interpreted, because M_0 is singular) is

$$\{A, B, C\colon ABA^{-1}B^{-1} = C, C^k = 1\}.$$

A moment's reflection will show that one must impose special conditions on k in order for the singularity of M_0 to resolve itself in M_1 and M_2. The condition is that any element of $\pi_1(M_0)$ corresponding to a singular point of M_0 should act freely on

the cosets of G/H_i, $i = 1, 2$. For the group G given above, one verifies that the condition is that $k = 7$.

We now seek two matrices A and B in $SL(3, Z/2)$ which have the properties that:

(i) A and B generate G.

(ii) $ABA^{-1}B^{-1}$ is of order 7.

One must be somewhat clever to find such matrices, but they do exist. One choice is:

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

To construct our surfaces M_1 and M_2, we break open our torus M_0 to get the six-sided polygon shown in FIGURE 1. The reader can check easily that when one identifies the sides corresponding to the same letters one obtains a torus with one singular point.

Now make seven copies of this polygon and label them according to the cosets of H_1 in G. A convenient way of doing this is by identifying G/H_1 with column vectors which are 0 or 1, but not all 0, by the identification

$$g \cdot H_1 \to g \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

We then glue the polygon labeled $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ with the polygon labeled $A\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ along the edges corresponding to the foot and head, respectively, of the arrow labeled A in FIGURE 1, and similarly for B and $C = ABA^{-1}B^{-1}$. When we are done, we obtain the gluing diagram for M_1 shown in FIGURE 2.

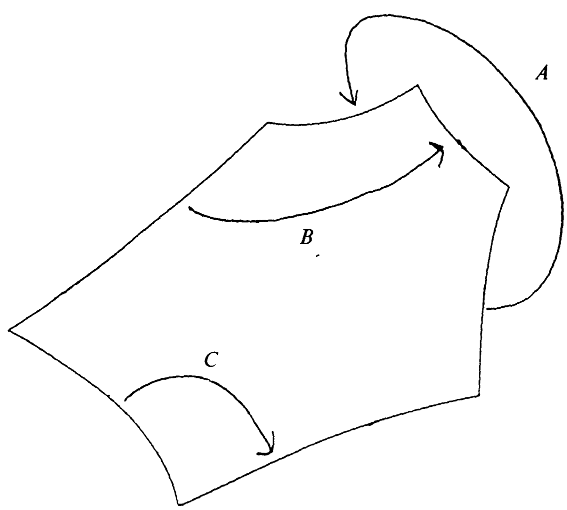


FIG. 1

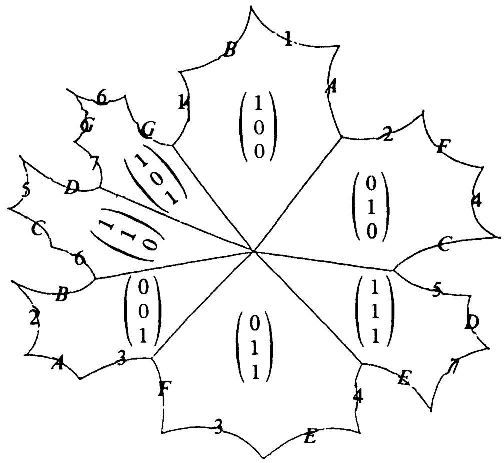


FIG. 2

We now do a similar construction to obtain M_2, but here we identify the cosets G/H_2 with column vectors, where a matrix $g \in G$ acts on a column vector by

$$(x, y, z) \to (x, y, z)(g^t)^{-1},$$

where g^t is the transpose of g.

Rather than find the inverses of A^t, B^t, and C^t, it turns out to be convenient to merely find the transposes of A, B, and C, and simply reverse the senses of the arrows in FIGURE 1. In this way, we obtain the gluing diagram for M_2 shown in FIGURE 3.

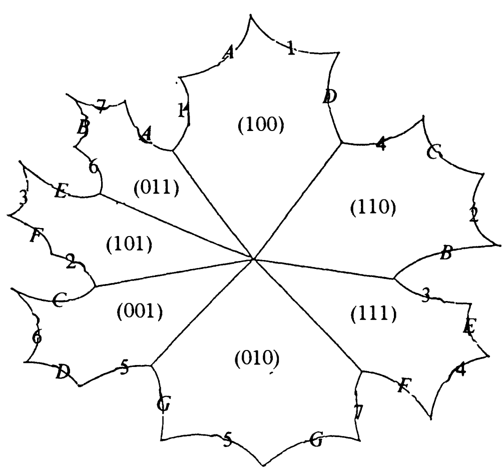


FIG. 3

To see that these surfaces are not isometric, notice that there is a short geodesic joining the two sides labeled 1, whose length is different from any other geodesic. Also, in FIGURE 2 there is a geodesic which joins the two pairs labeled A and B,

which again has length different from any other geodesic, and an analogous geodesic in FIGURE 3 joining the sides labelled E and F, which has the same length as the geodesic in FIGURE 2. In FIGURE 2, these two geodesics cross, but in FIGURE 3 they do not. Hence M_1 is not isometric to M_2.

It is difficult to see directly that these surfaces are isospectral. However, one may see readily that the set of lengths of closed geodesics is the same for both, by drawing closed loops on one and looking for the analogous loops on the other.

A somewhat simpler, but completely analogous, construction gives the genus 3 examples mentioned above. Here we begin with a sphere with three singular points, rather than a torus with one singular point, and the problem in group theory is to find matrices A and B in $SL(3,\mathbb{Z}/2)$ such that:

(i) A and B generate $SL(3,\mathbb{Z}/2)$

(ii) A, B, and AB are all of order 7.

One such choice is:

$$A=\begin{pmatrix}1&0&1\\1&1&0\\0&1&0\end{pmatrix}\qquad B=\begin{pmatrix}0&0&1\\1&0&0\\1&1&0\end{pmatrix}.$$

The resulting gluing diagrams are shown in FIGURES 4 and 5.

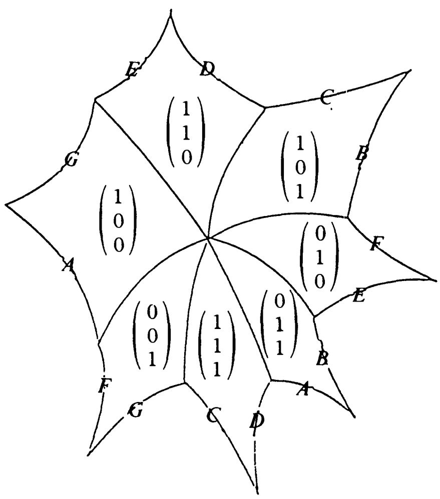


FIG. 4

It turns out that if one starts with a metric on the sphere with three singular points which is too symmetric (for instance, the unique metric of constant curvature -1), then M_1 will be isometric to M_2. To see this, one may take the gluing diagram of FIGURE 4 and reflect it in a mirror. After relabelling the letters in a suitable way, one sees that FIGURE 4 becomes the same as FIGURE 5. However, if the metric on the sphere is sufficiently bumpy, it will not be respected by reflection in the mirror, and M_1 and M_2 will not be isometric.

It is evident that this and similar constructions are mechanical, and I have found it amusing and simple to carry out some of these constructions with paper and scissors. Here is my favorite recipe to do this—the reader may enjoy trying his own:

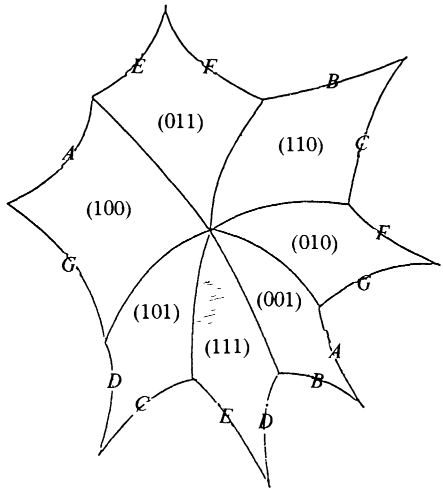


FIG. 5

Let us begin with the group $(Z/8) \times (Z/8)^*$ of Example 1. Here we may identify $G/H_1 = G/H_2 = Z/8$, by noting that $(a, b) \cdot (a, 0) = (1, ab)$, $ab \in Z/8$.

Now we consider the three-pronged figure of FIGURE 6. Each prong is identified with a generator of G (notice that G cannot be generated by two elements). For ease in pasting, we have taken each generator to be of order 2.

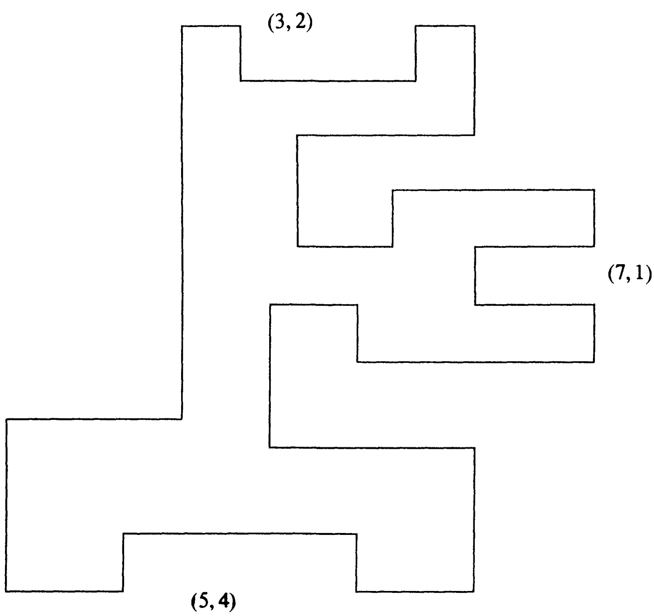


FIG. 6

Now make eight copies of FIGURE 6, and label them from 0 to 7, according to the cosets of G/H_1. Then glue them together according to the action of the generators

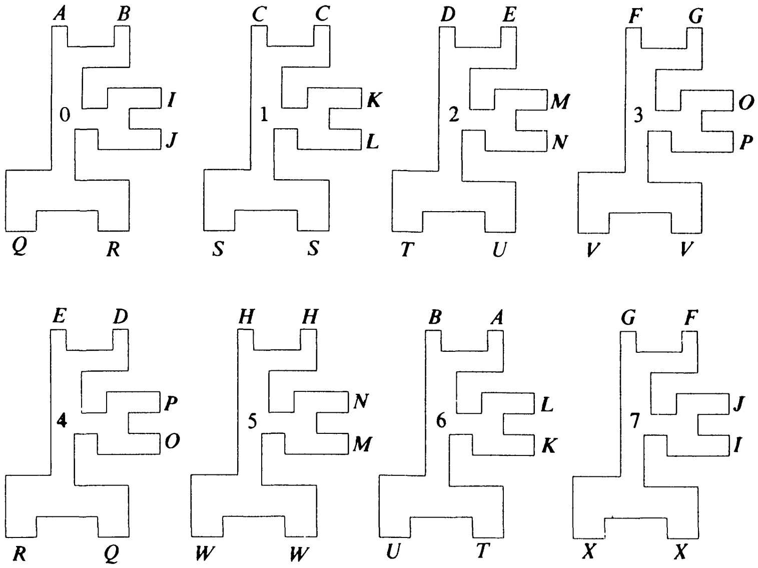


FIG. 7

on these cosets. Instructions for doing this are contained in FIGURE 7. Finally, repeat this with eight more copies, now using the gluing instructions from FIGURE 8, to obtain a second figure isospectral to the first.

For some of the fine points of paper constructions, see [5].

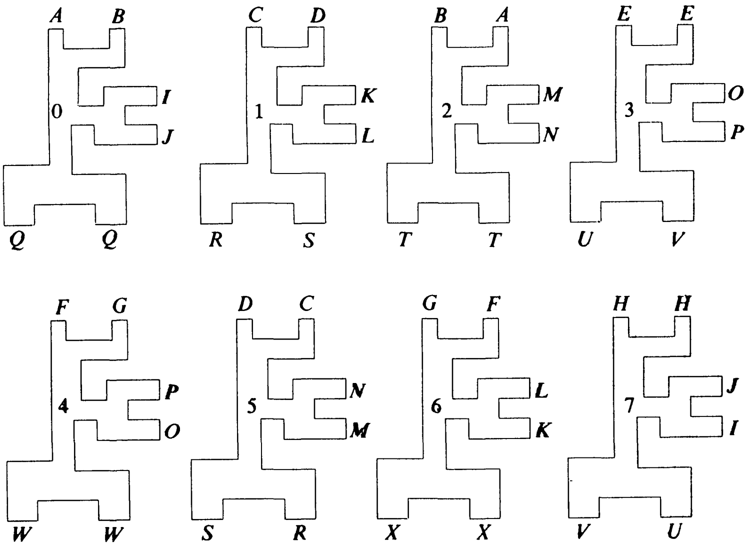


FIG. 8

5. Isospectral Deformations

A natural question arises from our discussion of § 4. We saw how one can construct examples of two manifolds M_1 and M_2 which are isospectral but not

isometric. Indeed, using Example 4 of § 3, one may find arbitrarily many manifolds $M_1, \ldots, M_n$ all of which are isospectral but none of which are isometric. Is it possible to find an infinite family M_t, for t ranging over the real numbers, such that all the M_t's are isospectral, but none of them are isometric?

To understand the extra challenge of this question, we observe that a theorem of Guillemin and Kazhdan [**11**] asserts that this can never happen when the M_t's have negative curvature. On the other hand, the construction of § 4 works perfectly well with manifolds of negative curvature, and indeed this is the easiest case, for then the fundamental groups are complicated enough to do the construction of § 4.

The problem of finding such isospectral families was solved by Gordon and Wilson in [**10**], and indeed their work came before the work of Sunada [**20**]. But it is not difficult to understand what they did in the framework we have set up. One may think of their construction as an analogue of Sunada's theorem for infinite groups.

Recall that the group-theoretic part of the argument of § 4 was to construct groups G, H_1 and H_2 satisfying the condition (†), such that H_1 was not conjugate to H_2 in G. One way to realize condition (†) is to take an automorphism φ of G which takes H_1 to H_2, such that φ takes each element h in H_1 to an element $\varphi(h)$ which is conjugate to h, but in such a way that the element of G which conjugates h to $\varphi(h)$ varies as h runs through H_1. Such an automorphism we will call an almost-inner automorphism of G (an inner automorphism being one which is conjugated by a fixed element of G).

Now, certain Lie groups admit families φ_t of such almost-inner automorphisms, and it was the observation of Gordon and Wilson that such families could be used to construct isospectral families. Perhaps the simplest such Lie group is

$$\left\{\begin{pmatrix} 1 & x_1 & x_2 & z_1 & 0 & 0 & 0 \\ 0 & 1 & 0 & y_1 & 0 & 0 & 0 \\ 0 & 0 & 1 & y_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & x_2 & z_2 \\ 0 & 0 & 0 & 0 & 0 & 1 & y_2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} : x_1, x_2, y_1, y_2, z_1, z_2 \in \mathbb{R}\right\}$$

with almost-inner automorphism

$$\varphi_t(x_1, x_2, y_1, y_2, z_1, z_2) = \begin{pmatrix} 1 & x_1 & x_2 & z_1 & 0 & 0 & 0 \\ 0 & 1 & 0 & y_1 & 0 & 0 & 0 \\ 0 & 0 & 1 & y_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & x_2 & z_2 + tx_1 \\ 0 & 0 & 0 & 0 & 0 & 1 & y_2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

The construction of isospectral families from this example of an almost-inner automorphism of G is carried out in detail in [**8**], to which we refer the reader for details.

6. The Future

It happens occasionally in mathematics that a problem that has vexed mathematicians for decades is so completely understood that it is difficult to look to the future. The present circumstance is one of those occasions. It appears that using some variation of these ideas allows one to find relatively easily examples that were difficult to imagine in the past (see, for instance, [3] or [19] for examples of this). Here, however, we will make the attempt to guess what may be profitable lines of research for the future.

A natural question is whether the techniques given here exhaust the methods of finding isospectral manifolds. Historical experience suggests that the answer to this question should be "no," but perhaps in sufficiently well-defined cases the answer may be "yes." For instance:

Question: Suppose S_1 and S_2 are two isospectral Riemann surfaces. Are there Riemann surfaces S' and S_0' such that S_1 and S_2 are sandwiched between them, as in the Sunada construction? We remark that the analogous problem in the number-theoretic case was settled in the affirmative by Robert Perlis [25].

Question: We constructed in § 4 two isospectral Riemann surfaces of genus 4. Are there examples of genus 2 or 3?

This question really has two parts: the first part asks if the group-theoretic construction given here for genus 4 can be made to work in lower genus. One can write down a finite (but long) list of all the possibilities, and in principle one can check each possibility individually. There are a few cases that almost work, but many of the cases are easily shown not to work. The bulk of the cases are unchecked. The second part of the question is to ask whether in genus 2 or 3 there might be other constructions which will work. Perhaps one can show that there are no isospectral Riemann surfaces of genus 2. See [13] for some ideas which tend in this direction, and which may or may not be able to be pushed forward.

A direction which promises to bear fruit in the future is the study of the small eigenvalues, such as λ_1. A beautiful characterization of λ_1 in terms of geometry was found by Cheeger in [7]. One believes that the behavior of λ_1 can play an interesting role in studying groups and manifolds which arise in number theory. But a complete elaboration of even the interesting questions of this line of thought is outside the scope of this paper.

A major message of the work described in this paper is the impact of finite group theory on fundamental problems in geometry and analysis. Since finite group theory is so well-developed, but in language quite far from the practicing geometer, it makes sense to ask:

Question: What do they know that we don't?

More precisely:

Question: What do they know that we are afraid to ask?

One cannot help but conclude with the observation that the techniques given here do not work for domains in the plane $\mathbb{R}^2$. So we can only repeat the question of Mark Kac:

Question: Can you hear the shape of a drum?

REFERENCES

1. P. Bérard and M. Berger, Le Spectre d'une Variété Riemannienne en 1982, in Spectra of Riemannian Manifolds, Kaigai Publications, 1983.
2. M. Berger, P. Gauduchon, and E. Mazet, Le Spectre d'une Variété Riemannienne, Springer Lecture Notes 194, 1971.
3. R. Brooks, Manifolds of negative curvature with isospectral potentials, *Topology*, 26 (1987) 63–66.
4. R. Brooks and R. Tse, Isospectral surfaces of small genus, *Nagoya Math J.*, 107 (1987) 13–24.
5. P. Buser, Isospectral Riemann surfaces, *Ann. Inst. Fourier* XXXVI, (1986), pp. 167–192.
6. I. Chavel, Eigenvalues in Riemannian Geometry, Academic Press, 1984.
7. J. Cheeger, A lower bound for the smallest eigenvalue of the Laplacian, In Gunning, Problems in Analysis (1970), pp. 195–199.
8. D. De Turck and C. Gordon, Isospectral deformations I: Riemannian structures on two-step nilspaces, *CPAM*, XL (1987) 367–387.
9. J. Dodziuk, Eigenvalues of the Laplacian and the heat equation, *Amer. Math Monthly*, 88 (1981) pp. 686–695.
10. C. Gordon and E. Wilson, Isospectral deformation of compact solvmanifolds, *J. Diff. Geom.*, 19 (1984) pp. 241–256.
11. V. Guillemin and D. Kazhdan, Some inverse spectral results for negatively curved 2-manifolds, *Topology*, 19 (1980) pp. 153–180.
12. R. Guralnick, Subgroups inducing the same permutation representation, *J. Alg.*, 81 (1983) pp. 312–319.
13. A. Haas, Length spectra as moduli for hyperbolic surfaces, *Duke Math. J.*, 5 (1985) pp. 923–935.
14. A. Ikeda, On lens spaces which are isospectral but not isometric, *Ann. Scient. E. Norm. Sup.*, 13 (1980) pp. 303–315.
15. M. Kac, Can one hear the shape of a drum? *Amer. Math. Monthly*, 73 (1966) pp. 1–23.
16. H. P. McKean, The Selberg trace formula as applied to a compact Riemann surface, CPAM, XXV (1972) pp. 225–246.
17. J. Milnor, Eigenvalues of the Laplace operator on certain manifolds, *Proc. NAS U.S.A.*, 51 (1964) p. 542.
18. A. Selberg, Harmonic analysis and discontinuous groups in weakly symmetric spaces with applications to Dirichlet series, *J. Ind. Math. Soc.*, 20 (1956) pp. 47–87.
19. T. Sunada, Gelfand's problem on unitary representations associated with discrete subgroups of PSL $(2,\mathbb{R})$, *Bull AMS.*, 12 (1985) pp. 237–238.
20. ______, Riemannian coverings and isospectral manifolds, *Ann. Math.*, 121 (1985) pp. 169–186.
21. M. F. Vignéras, Variétés Riemanniennes isospectrales et non isometriques, *Ann. Math.* 112, (1980) pp. 21–32.
22. H. Weyl, Über die Asymptotische Verteilung der Eigenwerte, *Nachr-der Königl. Ges d. Wiss. Zu Gottingen*, (1911) pp. 110–117.
23. R. S. Millman, Manifolds with the same spectrum, *Amer. Math. Monthly*, 90 (1983) pp. 553–555.
24. F. Gassmann, Bemerkung zu der vorstehenden Arbeit von Hurwitz, *Math. Z.*, 25 (1926) pp. 124–143.
25. R. Perlis, On the equation $\zeta_K(s) = \zeta_{K'}(s)$, *Journal of Number Theory*, 9 (1977) pp. 342–360.

The Editor's Corner: n Coins in a Fountain

ANDREW M. ODLYZKO* AND HERBERT S. WILF**

In Richard Guy's article in last month's Monthly **[6]** there appeared a number of elegant questions, one of which we will answer here.

An (n, k) *fountain* is an arrangement of n coins in rows such that there are exactly k coins in the bottom row, and such that each coin in a higher row touches exactly two coins in the next lower row. In FIG. 1 below we show a (28, 12) fountain. Let $f(n, k)$ be the number of (n, k) fountains, and let $f(n) = \sum_k f(n, k)$ be the number of fountains of n coins.

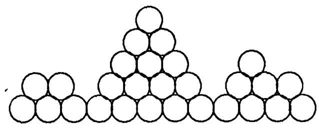


FIG. 1. A (28, 12) fountain.

The question of determining $\{f(n)\}$ was raised in [6], and attributed to J. G. Propp. The values of $f(n)$ for $1 \leqslant n \leqslant 18$ are 1, 1, 2, 3, 5, 9, 15, 26, 45, 78, 135, 234, 406, 704, 1222, 2120, 3679, 6385.

We will show that $\{f(n)\}$ has the remarkable generating function

$$F(x) = 1 + \sum_{n \geqslant 1} f(n) x^n = \cfrac{1}{1 - \cfrac{x}{1 - \cfrac{x^2}{1 - \cfrac{x^3}{\ddots}}}}. \tag{1}$$

This continued fraction was first studied by Ramanujan (see **[1]**, **[2]**, **[4]**, **[7]**). That the power series coefficients of this fraction have a combinatorial meaning is not new. Already in 1968 Szekeres **[8]** found such an interpretation and others were found later by Flajolet **[5]**. However the interpretation as the number of n-fountains seems particularly attractive, and it lends itself to a very transparent proof.

We say that an (n, k) fountain is *primitive* if its next-to-bottom row contains no empty positions (i.e., contains $k - 1$ coins), and let $g(n, k)$ be the number of primitive (n, k) fountains. By peeling off the bottom row of a primitive fountain we see that

$$f(n-k, k-1) = g(n, k) \qquad (n \geqslant k;\ k \geqslant 1). \tag{2}$$

Next, let $\mathscr{F}$ be an arbitrary (n, k) fountain, and suppose that the first empty position in the next-to-bottom row is the rth $(1 \leqslant r \leqslant k - 1)$ (in FIG. 2 we have

*AT & T Bell Laboratories, Murray Hill, NJ 07974.
**Supported in part by the United States Office of Naval Research.

$r = 3$). Then by drawing a dotted line to the right of the rth coin in the bottom row, as in Fig. 2, we split $\mathscr{F}$ into an (n', r) primitive fountain and an $(n - n', k - r)$ not-necessarily-primitive fountain.

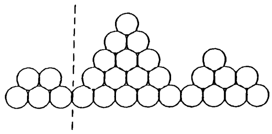


FIG. 2. Split off a primitive fountain.

The factorization is unique, and so we have

$$f(n,k) = \sum_{n', r \geq 0} g(n', r) f(n - n', k - r) \tag{3}$$

for all $n, k \geq 1$, where $f(0,0) = 1$ and $g(0,0) = 0$.

Hence if $F(x, y)$ and $G(x, y)$ are the ordinary generating functions of the sequences f, g respectively, then from (3) we find that $F = 1 + FG$ and from (2) that $G(x, y) = xyF(x, xy)$, whence

$$\begin{aligned} F(x,y) &= \frac{1}{1 - xyF(x, xy)} \\ &= \cfrac{1}{1 - \cfrac{xy}{1 - x^2yF(x, x^2y)}} \\ &= \cdots \\ &= \cfrac{1}{1 - \cfrac{xy}{1 - \cfrac{x^2y}{1 - \cfrac{x^3y}{\ddots}}}}. \end{aligned} \tag{4}$$

In particular, if we put $y = 1$ we obtain the generating function for the numbers of n-fountains in the form (1), as claimed.

Remarks

1. An easier problem would have resulted if we had asked for $h(k) = \Sigma_n f(n, k)$, the number of fountains whose bottom row contains k coins, without regard to the total number of coins. If we put $x = 1$ in (4) we find that $\{h(k)\}$ is generated by $\mathrm{H}(y)$ where $\mathrm{H}(y) - y\mathrm{H}^2(y) = 1$. Hence the $\{h(k)\}$ are the Catalan numbers.

2. The continued fraction (1) of Ramanujan can be rewritten in various forms, for instance as

$$F(x) = \frac{\displaystyle\sum_{r \geq 0} (-1)^r \frac{x^{r(r+1)}}{(1-x)(1-x^2)\cdots(1-x^r)}}{\displaystyle\sum_{r \geq 0} (-1)^r \frac{x^{r^2}}{(1-x)(1-x^2)\cdots(1-x^r)}}. \tag{5}$$

We will now use this expression for $F(x)$ to deduce the asymptotics of $f(n)$. More precisely, we will show that

$$f(n) = cx_0^{-n} + O\left((5/3)^n\right) \qquad (n \to \infty) \tag{6}$$

where $c = 0.312363324596741\ldots$ and $x_0 = 0.576148769142756\ldots$.

Let $p(x)$ and $q(x)$ denote, respectively, the numerator and the denominator on the right side of (5), and let $p_n(x)$, $q_n(x)$ denote the nth partial sums of the series for $p(x)$ and $q(x)$. It is clear that p and q are analytic in $|x| < 1$, so F is meromorphic there. We will show that $q(x)$ has a simple, real zero $x_0 \in (0.57, 0.58)$, and no other zeros in the disk $|x| < 0.62$, while $p(x_0) > 0$. It will then follow that (6) holds, and more precise numerical calculations yield the more accurate values of x_0 and $c = -p(x_0)/(x_0 q'(x_0))$ stated above.

The proof of our claim about x_0 can be obtained fairly easily using numerical methods to compute multiplicities of zeros of analytic functions. Here we sketch a proof that can be carried out by hand.

Write $q_3(x) = a(x)/((1-x)(1-x^2)(1-x^3))$ where

$$a(x) = 1 - 2x - x^2 + x^3 + 3x^4 + x^5 - 2x^6 - x^7 - x^9,$$

and consider $b(x) = \prod_{j=1}^{9}(x - x_j)$, where the x_j are 0.57577, $-0.46997 \pm i0.81792$, $0.74833 \pm i0.07523$, $-1.05926 \pm i0.36718$, $0.49301 \pm i1.58185$, in that order (these x_j's are approximations to the zeros of $a(x)$). An easy hand calculation shows that if $\{a_k\}$ and $\{b_k\}$ are the coefficients of $a(x)$ and $b(x)$ respectively, then $\sum |a_k - b_k| \leqslant 1.7 \times 10^{-4}$, and so $|a(x) - b(x)| \leqslant 1.7 \times 10^{-4}$ for all $|x| \leqslant 1$. Another such calculation shows that $|b(x)| \geqslant 8 \times 10^{-4}$ for all $|x| = 0.62$.

On the other hand, for $u \in (0, 0.62)$ and $|x| = u$ we have

$$\left|\frac{x^{(k+1)^2-k^2}}{1-x^{k+1}}\right| \leqslant \frac{u^{2k+1}}{1-u^{k+1}} \leqslant \frac{u^9}{1-u^5} \qquad (k \geqslant 5)$$

and, therefore,

$$\left|\sum_{k=4}^{\infty}(-1)^k \frac{x^{k^2}}{\prod_{j=4}^{k}(1-x^j)}\right| \leqslant \frac{u^{16}}{1-u^4}\sum_{m\geqslant 0}\left(\frac{u^9}{1-u^5}\right)^m \leqslant 6 \times 10^{-4}.$$

Therefore, by Rouché's theorem, $q(x)$ and $b(x)$ have the same number of zeros in $|x| \leqslant 0.62$, namely, 1. Further, since q has real coefficients, its zero is real, and we call it x_0. A brief calculation shows that $x_0 \in (0.57, 0.58)$.

Finally, we must show that $p(x_0) > 0$. However, since the successive summands in the definition of $p(x)$ decrease in absolute magnitude for each fixed real x, $0 < x < 0.6$, we have $p(x) > 1 - x^2/(1-x)$, which is positive for $0 < x < 0.6$, and the claim is established.

We note that the asymptotic expansion that we have proved is very accurate, even for moderately small values of n. For instance, $f(120) = 1.700213368\ldots \times 10^{28}$, while $f(120) - cx_0^{-120} = 1.59\ldots \times 10^9$.

3. To make the correspondence between our problem and the one that Szekeres solved quite explicit, we draw the slant lines in Fig. 1 that are shown in FIG. 3, and let x_i denote the number of coins on the ith slant line. If we let $q_i = x_i + i - 1$ $(i = 1, k)$ then we see that $f(n, k)$ counts partitions with a fixed 'rate of climb.'

Precisely, $f(n, k)$ is the number of partitions of the integer $n + \binom{k}{2}$, whose largest part is k and whose ith-from-smallest part is $\geqslant i$ for each $i = 1, k$, and that is exactly the problem that Szekeres treated in [8].

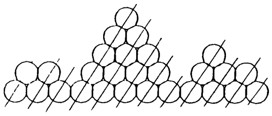


FIG. 3. Draw the slant lines.

4. A problem that is similar to this one has been considered by Auluck [3]. He dealt with arrangements of coins that satisfy our conditions and are further subject to the condition that the set of coins in each horizontal row forms a single contiguous block.

REFERENCES

1. George E. Andrews, An introduction to Ramanujan's 'lost notebook,' this MONTHLY, 86 (1979) 89–108.
2. ______, The theory of partitions, Encycl. Math. and its Appl., vol. 2, Addison-Wesley, 1976.
3. F. C. Auluck, On some new types of partitions associated with generalized Ferrers graphs, *Proc. Cambr. Phil. Soc.*, 47 (1951) 679–686.
4. S. Bhargava and Chandrasekhar Adiga, On some continued fraction identities of Srinavasa Ramanujan, *Proc. Amer. Math. Soc.*, 92 (1984) 13–18.
5. P. Flajolet, Combinatorial aspects of continued fractions, *Discrete Mathematics*, 32 (1980) 125–161.
6. Richard Guy, The strong law of small numbers, this MONTHLY (October, 1988).
7. M. D. Hirschhorn, A continued fraction of Ramanujan, *J. Austral. Math. Soc.* (Ser. A) 29 (1980) 80–86.
8. G. Szekeres, A combinatorial interpretation of Ramanujan's continued fraction, *Canadian Math. Bull.*, 11 (1968) 405–408.

LETTERS TO THE EDITOR

Editor:

I was appalled to read C. Smorynski's vicious attack on James Henle's elegant and beautiful book, *An Outline of Set Theory*, and was amazed that the editors of this journal let such a vituperative review be printed—the MONTHLY has an enormous circulation, and its reviews are highly influential. To allow this fine book to be excoriated in front of such a large audience is extremely unfortunate.

The gist of Smorynski's diatribe is that Henle did not write the sort of book which Smorynski would like to see written. It is as if an orange is being accused of not being an apple.

1. Henle is accused of not writing a text organized on historical lines—he didn't set out to write one.

2. Henle is accused of writing a book in which mathematics is motivated by having the students actually do it, rather than passively consuming it. That's just what he wanted to do.

3. Preposterously, Henle is accused of being a mere problem solver. I find this charge incredible, since the problems are designed exactly to lead students to the basics of modern set theory, and to its connection with the rest of mathematics. This is not problem solving for its own sake!

4. Henle is accused of not including enough material. But he didn't set out to write a comprehensive text. He wrote an introduction to *undergraduate* set theory. In fact, Henle covers far more material than any other book I know of on this level, even covering topics that are often missing in beginning graduate texts.

5. Henle's history is wrong. Smorynski may be right on this, but he should cast no stones, since his own history is wrong: it was not Zermelo who formalized the hierarchy of sets; the process of doing so was a long one, beginning with Mirimanov in 1917 and ending with von Neumann in 1925. With all due respect to the importance of the history of mathematics, it is astounding that the bulk of Smorynski's comments concerns a few paragraphs in a 135 page book whose major focus is elsewhere.

If Smorynski would like to write a comprehensive set theory text organized on historical lines, he is welcome to write one. He should not condemn others for writing different books.

I encourage the reader interested in an undergraduate set theory text to consider Henle's book (which, by the way, is designed so that it can be used in a conventional class as well as in a Moore-method class). In fact, Henle's book is so well organized and written that it would make a pleasurable introduction for anyone wanting to learn basic set theory. For more detailed comments, see my own review in the *Journal of Symbolic Logic*, vol. 52 No. 4, December 1987, p. 1048.

I also urge the editorial board of the MONTHLY to carefully consider their editorial policy on book reviews. Polemical diatribes are appropriate in letters to the editor. They are not appropriate in reviews.

Sincerely,

JUDITH ROITMAN
Professor of Mathematics
University of Kansas

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

Conjectured Isoptic Characterization of a Circle

M. S. KLAMKIN
Department of Mathematics, University of Alberta, Edmonton, Alberta, Canada T6G 2G1

The **isoptic curve** of a given smooth convex region C is the locus of the intersections of pairs of tangents to C which meet at a constant angle, say $\pi - \beta$. It is known that if the isoptic curve is a circle, the region C need not be a circle. For example, it can be an ellipse, a hyperbola or a deltoid [**1**]. However, Green [**2**] has shown that if β is an irrational multiple of π or if $\beta = (m/n)\pi$ where m/n in its lowest terms has an even numerator, then C must be a circle concentric with the isoptic locus. Nevertheless, it is conjectured, that if two isoptic curves of the region C are concentric circles, then C is also a circle. Prove or disprove. In view of the result of Green, it is assumed that the two angles β_1 and β_2 are each of the form $m\pi/n$ with m odd.

REFERENCES

1. R. C. Yates, A Handbook on Curves and their Properties, J. W. Edwards, Ann Arbor, 1947, pp. 138–140.
2. J. W. Green, Sets subtending a constant angle on a circle, *Duke Math. J.*, 17(1950) 263–267.

A Conjecture Concerning Magic Squares

J. DÉNES
Csaba Utca 10 V 42, 1122 Budapest, Hungary

A. D. KEEDWELL
Department of Mathematics, University of Surrey, Guildford, Surrey, England GU2 5XH

A **magic square** of order n is an $n \times n$ matrix whose entries are n^2 usually distinct, and usually consecutive, integers so arranged that the sums of the entries in each row, each column and each main diagonal of the matrix are equal. A classical method of constructing such squares is by means of pairs of orthogonal latin squares. See, for example, [**6**] or, for a more recent account, Chapter 6 of [**4**]. By a **latin square** of order n is meant an $n \times n$ matrix containing n different symbols

each repeated n times in such a way that each symbol occurs just once in each row and column of the square. The definition of **orthogonal** is given in the next paragraph. If L_1 and L_2 are two orthogonal latin squares of order n with suitable diagonal properties and having appropriately chosen integers as entries, then the matrix $L_1 + L_2$ is a magic square. The question which we wish to raise is whether it is true conversely that every magic square can be represented as a sum of two latin squares. We provide some evidence for the truth of our conjecture and pose two alternative weaker conjectures.

Let M be an $n \times n$ magic square whose entries are the consecutive integers $0, 1, \ldots, n^2 - 1$ and let us write these integers as numbers to base n. Then the first digits of each such number define a square L_1, the second digits form a second square L_2 and we have $M = nL_1 + L_2$. Both nL_1 and L_2 are **equi-n-squares**: that is, each involves just n symbols each repeated n times but they need not be latin. However, if M is to have entries which are consecutive integers, the squares L_1 and L_2 must be **orthogonal**: that is, if they are superimposed, each of the n different integers which occur as the entries of one square must occur in the same cell with each of the n different integers which are the entries of the other square exactly once. (Of course, the squares L_1 and L_2 presently being discussed both have the same set $0, 1, \ldots, n - 1$ of integers as entries but later in this paper this will not always be the case.)

Also, if S_1 denotes the sum of the elements of one of the rows, columns or main diagonals of L_1 and S_2 denotes the sum of the elements of the same row, column or main diagonal of L_2, then the diophantine equation $nS_1 + S_2 = \frac{1}{2}n(n^2 - 1)$ must be satisfied for that row, column or main diagonal. To see this, we note that the sum of the integers $0, 1, \ldots, n^2 - 1$ is $\frac{1}{2}n^2(n^2 - 1)$. So if each row of the magic square is to have the same sum, that sum must be $\frac{1}{2}n(n^2 - 1)$. The common sum of each row, column and diagonal of a magic square is usually called its **magic sum**. E. Stern [7] has pointed out that these conditions are sufficient for L_1 and L_2 to define a magic square. He has also pointed out that, if L_1 and L_2 are themselves magic (in which case, $S_1 = S_2 = \frac{1}{2}n(n - 1)$ for each row, column and main diagonal), then the last condition is automatically fulfilled.

In example 1 below, L_1 and L_2 are both latin and magic and then $M = nL_1 + L_2$ is certainly a sum of two latin squares.

Example 1. $(n = 5)$

$$M = \begin{array}{|ccccc|}\hline 2 & 8 & 24 & 16 & 10 \\ 5 & 22 & 18 & 14 & 1 \\ 21 & 15 & 12 & 3 & 9 \\ 19 & 11 & 0 & 7 & 23 \\ 13 & 4 & 6 & 20 & 17 \\ \hline\end{array} \qquad L_1 = \begin{array}{|ccccc|}\hline 0 & 1 & 4 & 3 & 2 \\ 1 & 4 & 3 & 2 & 0 \\ 4 & 3 & 2 & 0 & 1 \\ 3 & 2 & 0 & 1 & 4 \\ 2 & 0 & 1 & 4 & 3 \\ \hline\end{array} \qquad L_2 = \begin{array}{|ccccc|}\hline 2 & 3 & 4 & 1 & 0 \\ 0 & 2 & 3 & 4 & 1 \\ 1 & 0 & 2 & 3 & 4 \\ 4 & 1 & 0 & 2 & 3 \\ 3 & 4 & 1 & 0 & 2 \\ \hline\end{array}$$

In [3], J. Chernick has shown that every magic square of order 3, whether involving consecutive integers or not, can be expressed in terms of three variables k, a, b as in Figure 1 and that every magic square of order 4, again involving not-necessarily-consecutive integers, can be expressed in the form shown in Figure 2 which involves eight variables. The latter form had been obtained earlier by E. Bergholt [1] but without a satisfactory proof that it was universal.

$k+a$	$k-a-b$	$k+b$
$k-a+b$	k	$k+a-b$
$k-b$	$k+a+b$	$k-a$

FIGURE 1.

$A-a$	$C+a+c$	$B+b-c$	$D-b$
$D+a-d$	B	C	$A-a+d$
$C-b+d$	A	D	$B+b-d$
$B+b$	$D-a-c$	$A-b+c$	$C+a$

FIGURE 2.

From Figure 1, it is easy to see that every magic square of order 3 is the sum of two orthogonal latin squares in many ways. Two such decompositions are shown in Figure 3. When the square has the consecutive integers 0 to 8 as its entries, we necessarily have $k = 4$ (since the magic sum $3k$ is 12) and so $a + b = 4$, implying $a = 3$ and $b = 1$ (since $a \neq b$). Then the decomposition of Figure 3(b) gives the form $M = nL_1 + L_2$.

$$\begin{array}{|ccc|}\hline k+a & k-a & k \\ k-a & k & k+a \\ k & k+a & k-a \\ \hline\end{array} + \begin{array}{|ccc|}\hline 0 & -b & b \\ b & 0 & -b \\ -b & b & 0 \\ \hline\end{array}$$

FIGURE 3(a).

$$\begin{array}{|ccc|}\hline \frac{3}{4}k+a & \frac{3}{4}k-a & \frac{3}{4}k \\ \frac{3}{4}k-a & \frac{3}{4}k & \frac{3}{4}k+a \\ \frac{3}{4}k & \frac{3}{4}k+a & \frac{3}{4}k-a \\ \hline\end{array} + \begin{array}{|ccc|}\hline \frac{1}{4}k & \frac{1}{4}k-b & \frac{1}{4}k+b \\ \frac{1}{4}k+b & \frac{1}{4}k & \frac{1}{4}k-b \\ \frac{1}{4}k-b & \frac{1}{4}k+b & \frac{1}{4}k \\ \hline\end{array}$$

FIGURE 3(b).

Bergholt has obtained a somewhat more restricted result for magic squares of order 4: namely, that if such a square has the extra property that the sums of the elements of each quarter are equal, then the magic square is the sum of two orthogonal latin squares. In Figure 2, the requirement that the elements of each quarter have the same sum requires that $d = a + c$. If we then eliminate the variable a and write $A + c = \alpha$, $B + b = \beta$, $C + d = \gamma$ and $D = \delta$, we get the form shown in Figure 4 which is clearly the sum of two orthogonal latin squares. Because any latin square of order 4 which has an orthogonal mate must be isotopic to Klein's four-group $\mathbb{Z}_2 \times \mathbb{Z}_2$ (see **[4]**, pages 34 and 129) and can only be magic if it has one of the three forms shown in Figure 5, it is clear that, conversely, a magic square of order 4 having consecutive integers as entries and constructed as the sum of two orthogonal latin squares of that order has equal quarters. Thus, for example, the magic square shown in Figure 6 cannot be expressed as the sum of two orthogonal latin squares but, despite this, it can be expressed as the sum of two orthogonal squares one of which is row-latin and the other column-latin, as shown. An equi-n-square is **row-latin** if each symbol occurs just once in each of its rows but not necessarily in each column. A **column-latin** square is similarly defined. The idea of constructing a magic square as the sum of a row-latin square and a column-latin square goes back at least as far as Ozanam: see pages 99 and 100 of **[6]** for magic squares of orders 6 and 8 constructed in this way.

$$\begin{array}{|cccc|}\hline \alpha-d & \gamma & \beta-c & \delta-b \\ \delta-c & \beta-b & \gamma-d & \alpha \\ \gamma-b & \alpha-c & \delta & \beta-d \\ \beta & \delta-d & \alpha-b & \gamma-c \\ \hline\end{array} = \begin{array}{|cccc|}\hline \alpha & \gamma & \beta & \delta \\ \delta & \beta & \gamma & \alpha \\ \gamma & \alpha & \delta & \beta \\ \beta & \delta & \alpha & \gamma \\ \hline\end{array} + \begin{array}{|cccc|}\hline -d & 0 & -c & -b \\ -c & -b & -d & 0 \\ -b & -c & 0 & -d \\ 0 & -d & -b & -c \\ \hline\end{array}$$

FIGURE 4.

$$\begin{array}{|cccc|}\hline 0 & 1 & 2 & 3 \\ 2 & 3 & 0 & 1 \\ 3 & 2 & 1 & 0 \\ 1 & 0 & 3 & 2 \\ \hline\end{array} \qquad \begin{array}{|cccc|}\hline 0 & 1 & 2 & 3 \\ 3 & 2 & 1 & 0 \\ 1 & 0 & 3 & 2 \\ 2 & 3 & 0 & 1 \\ \hline\end{array} \qquad \begin{array}{|cccc|}\hline 0 & 1 & 2 & 3 \\ 2 & 3 & 0 & 1 \\ 1 & 0 & 3 & 2 \\ 3 & 2 & 1 & 0 \\ \hline\end{array}$$

FIGURE 5.

$$\begin{array}{|cccc|}\hline 0 & 1 & 15 & 14 \\ 12 & 13 & 3 & 2 \\ 11 & 6 & 8 & 5 \\ 7 & 10 & 4 & 9 \\ \hline\end{array} = \begin{array}{|cccc|}\hline 0 & 0 & 12 & 12 \\ 12 & 12 & 0 & 0 \\ 8 & 4 & 8 & 4 \\ 4 & 8 & 4 & 8 \\ \hline\end{array} + \begin{array}{|cccc|}\hline 0 & 1 & 3 & 2 \\ 0 & 1 & 3 & 2 \\ 3 & 2 & 0 & 1 \\ 3 & 2 & 0 & 1 \\ \hline\end{array}$$

FIGURE 6.

The above results lead us to pose the following two questions (though we are not sure which is likely to prove the more difficult to resolve):

Problem 1. Can every magic square be expressed as the sum of two orthogonal squares one of which is row-latin and the other column-latin?

Problem 2. Can every magic square be expressed as the sum of two not-necessarily-orthogonal latin squares?

We draw the attention of the reader to some further results and questions which are relevant.

First, in [**3**], Chernick has given a form of expression valid for every magic square of order n, $n \geqslant 5$, showing that it depends on $n^2 - 2n$ arbitrary constants. (The entries of the square are not restricted to being consecutive integers.)

Second, we draw attention to a theorem of G. Birkhoff [**2**] concerning the representation of a doubly stochastic matrix as a sum of permutation matrices. A square matrix with nonnegative entries is called **doubly stochastic** if the sum of the elements of each of its rows and columns is equal to unity. Thus, if all the elements of a magic square are divided by the magic sum, the resulting matrix is doubly stochastic. Likewise, if all the elements of a latin square are divided by the sum of the elements of any one row (or column), the resulting matrix is doubly stochastic. Birkhoff's theorem states that any $n \times n$ doubly stochastic matrix D can be expressed in the form $D = \sum_{i=1}^{s}\theta_i P_i$, for some positive integer s, where $P_1, P_2, \ldots, P_s$ are $n \times n$ permutation matrices and $\theta_1, \theta_2, \ldots, \theta_s$ are nonnegative numbers whose sum is unity. An example of such a decomposition is given in Figure 7. A proof of the theorem can be found, for example, on page 36 of [**5**]. Let us note further that any doubly stochastic latin square L of order n can be expressed in the form $L = \sum_{i=1}^{n}\theta_i P_i$, where $P_1, P_2, \ldots, P_n$ are *disjoint* permutation matrices and $\sum_{i=1}^{n}\theta_i = 1$. Here, each different element of L defines one of the permutation matrices. An example is shown in Figure 8.

$$\begin{pmatrix} \frac{1}{4} & \frac{2}{3} & \frac{1}{12} \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{2} \\ \frac{7}{12} & 0 & \frac{5}{12} \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} + \frac{1}{6}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \frac{1}{12}\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

FIGURE 7.

$$\begin{pmatrix} 0 & \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} \end{pmatrix} = 0\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} + \frac{2}{3}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \frac{1}{3}\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

FIGURE 8.

These observations suggest the following question:

Problem 3. Is it possible to write every doubly stochastic magic square M as a weighted sum of doubly stochastic latin squares $L_1, L_2, \ldots, L_s$, say

$$M = \sum_{i=1}^{s} \theta_i L_i, \text{ where } \sum_{i=1}^{s} \theta_i = 1?$$

If the answer to Problem 3 is "YES", then we have solved a weaker version of Problem 2 above: namely, we have shown that every magic square can be expressed as a sum of latin squares.

Acknowledgements. (1) The research work of which this article forms a part was assisted by a grant from the Soros Foundation. (2) The authors wish to thank R. K. Guy for drawing their attention to the existence of the compilation "Sources in Recreational Mathematics" by D. Singmaster and kindly supplying us with copies of it.

REFERENCES

1. E. Bergholt, The magic square of sixteen cells: A new and completely general formula, *Nature*, 83(1910) 368–369.
2. G. Birkhoff, Three observations in linear algebra, (Spanish) *Univ. Nac. Tucumán. Revista A*, 5(1946) 147–151; MR**8**-561.
3. J. Chernick, Solution of the general magic square, this MONTHLY 45(1938) 172–175.
4. J. Dénes and A. D. Keedwell, Latin Squares and their Applications, Akadémiai Kiadó, Budapest/English Universities Press, London/Academic Press, New York, 1974.
5. H. Minc, Permanents, Addison-Wesley, Reading, 1978.
6. M. Ozanam, Recreations in Science and Natural Philosophy, Revised English edition by Edward Riddle, Thomas Tegg, London, 1844.
7. E. Stern, General formulas for the number of magic squares belonging to certain classes, this MONTHLY, 46(1939) 555–581.

Edited by Dennis DeTurck, David J. Hallenbeck, and Anita E. Solow

A Simple Cover-up Game

V. J. Baston and F. A. Bostock

Two decks, each with n cards numbered 1 to n, are shuffled separately, and the top card from each deck is laid face down on the table. One of two players now looks at the numbers on the cards and chooses to turn one of them face up. The other player selects one of the cards and receives from his opponent an amount equal to the number of his selected card. What is a fair entrance fee for the second player to play this game?

The problem can be modelled mathematically as a two-person, zero-sum game Ω_n as follows. Each element of an ordered pair (x_1, x_2) is chosen independently from $\{1, 2, \ldots, n\}$ with probability $1/n$. Only player 1 sees (x_1, x_2) and, after seeing it, he covers either x_1 or x_2 before showing it to player 2. Player 2 now either accepts the uncovered number or rejects it and chooses the covered one. The number so chosen is the payoff to player 2.

Note that the value of Ω_n to player 2 is then the entrance fee required.

Before presenting a solution of Ω_n we will find some preliminary discussion useful. We will denote by $E(X, Y)$ the expectation to player 2 when player 1 plays the strategy X and player 2 the strategy Y. Notice that player 2 is trying to maximize $E(X, Y)$ whilst player 1 is trying to minimize $E(X, Y)$. If X is a fixed strategy for player 1, then a best reply to X is any strategy Y^* such that $E(X, Y^*)$ is the maximum of $E(X, Y)$ over all strategies Y for player 2. Similarly, a best reply to a strategy Y of player 2 is any strategy X^* such that $E(X^*, Y)$ is the minimum of $E(X, Y)$ over all strategies X for player 1. Let X^* and Y^* be strategies for the players such that X^* is a best reply to Y^* and Y^* is a best reply to X^*. Then it is immediate from the definition of optimal strategy that X^* and Y^* are optimal. Technically, a solution of the game is obtained when an optimal strategy for each player has been found. Although a game can have many solutions, $E(X^*, Y^*)$ is the same whatever optimal strategies X^* and Y^* are used for the players, and it is referred to as the value of the game to player 2. The reader who is not familiar with the rudiments of elementary game theory is recommended to consult [3].

In playing Ω_n player 2 clearly has simpler decisions to make than player 1, so it is natural to first consider what strategies this player can adopt in order to do well. It is routine to calculate that the maximum and minimum of the two numbers seen by player 1 have expected values $(n + 1)(4n - 1)/6n$ and $(n + 1)(2n + 1)/6n$, respectively. Since player 2 knows he is seeing either the maximum or the minimum, a reasonable strategy for him to adopt is to accept the uncovered number if and only if it is at least $(n + 1)/2$. A consideration of cases in which n is small suggests that this strategy is indeed optimal and leads to the following theorem.

Theorem. *The following strategies are optimal in Ω_n. For player* 1, *cover a maximum number if and only if the sum of the numbers is at least* $2[(n + 2)/2]$ *where* $[x]$ *denotes the integral part of* x. *For player* 2 *accept the uncovered number if and*

only if it is at least $[(n+2)/2]$. *The value of* Ω_n *is* $(7n^2+6n+2)/12n$ *if* n *is even and* $(7n^2+6n-1)/12n$ *if* n *is odd.*

Proof. Let $d=[(n+2)/2]$ and denote the strategies in the statement of the theorem by X^* and Y^*, respectively. Let player 2 adopt Y^*. Then, for a best reply, player 1 should cover a maximum if both are at least d and cover a minimum if both are strictly less than d; in the other cases, he can cover either number. Hence it is easy to see that X^* is a best reply to Y^*.

Now let player 1 adopt X^*. Suppose player 2 sees m, where $m \geqslant d$. This could have arisen via the pairs (m, m), (m, t), and (t, m), where $m < t \leqslant n$ or $1 \leqslant t < 2d-m$. Hence if player 2 rejects m, he obtains m for the pair (m, m) and t for each of the pairs (m, t) and (t, m), where $m < t \leqslant n$ or $1 \leqslant t < 2d-m$; since there is a total of $4d-4m+2n-1$ such pairs (occurring with equal probability), player 2's conditional expectation would then be

$$\left(m+2\sum_{t=m+1}^{n} t+2\sum_{t=1}^{2d-m-1} t\right)\Bigg/(4d-4m+2n-1)$$
$$=(n^2+n+4d^2-4dm-2d+m)/(4d-4m+2n-1).$$

Thus he will not do worse by accepting m provided

$$m(4d-4m+2n-1) \geqslant n^2+n+4d^2-4dm-2d+m,$$

i.e.,

$$2(m-d)(n-2m+2d-1)+(2d-n-1)n \geqslant 0.$$

By considering separately the cases n odd and n even it is easy to check that all the brackets are nonnegative for $[(n+2)/2]=d \leqslant m \leqslant n$, so the inequality holds.

Now suppose player 2 sees m where $m < d$. This could have arisen via the pairs (m, m), (m, t) and (t, m), where $0 \leqslant t < m$ or $2d-m \leqslant t \leqslant n$. Hence if player 2 rejects m, he obtains m for the pair (m, m) and t for each of the pairs (m, t) and (t, m), where $0 \leqslant t < m$ or $2d-m \leqslant t \leqslant n$; since there is a total of $4m-4d+2n+1$ such pairs (occurring with equal probability), player 2's conditional expectation would then be

$$\left(m+2\sum_{t=1}^{m-1} t+2\sum_{t=2d-m}^{n} t\right)\Bigg/(4m-4d+2n+1)$$
$$=(n^2+n-4d^2+4dm+2d-m)/(4m-4d+2n+1).$$

Thus he will not do worse by rejecting m provided

$$n^2+n-4d^2+4dm+2d-m \geqslant m(4m-4d+2n+1),$$

i.e.,

$$2(d-m-1)(n+1-2d+2m)+(n+2)(n+1-2d)+4m \geqslant 0.$$

Since all the brackets are nonnegative for $0 \leqslant m < d=[(n+2)/2]$ the inequality holds.

It follows that Y^* is a best reply to X^*, whence X^* and Y^* are optimal.

We obtain the value of Ω_n by calculating $E(X^*, Y^*)$. Let player 1 adopt X^* then, by our previous analysis:

(i) player 2 will see the number $m \geqslant d$ with probability $(4d - 4m + 2n - 1)/n^2$ and, in this case, since he accepts m, there will be a contribution $m(4d - 4m + 2n - 1)/n^2$ to his expectation;

(ii) there will be contribution $(n^2 + n - 4d^2 + 4dm + 2d - m)/n^2$ to his expectation when player 2 rejects $m < d$. Hence

$$\begin{aligned} n^2E(X^*, Y^*) &= \sum_{m=d}^{n} m(4d - 4m + 2n - 1) \\ &\quad + \sum_{m=1}^{d-1} (n^2 + n - 4d^2 + 4dm + 2d - m) \\ &= \frac{1}{2}n(n+1)(6d - (2n+13)/3) - \frac{1}{2}d(d-1)(2n + (16d-8)/3) \end{aligned}$$

after some routine manipulation.

The result now follows easily.

It is natural to consider the corresponding game where there is only one deck of cards and the top two members are laid face down on the table. Trivial modifications in the proof of our theorem show that the players have the same optimal strategies as in the original game and that the value is $7(n+1)/12$ if n is odd and $(7n^2 - 4)/12(n-1)$ if n is even.

From a mathematical point of view it is also natural to consider the infinite game Ω corresponding to Ω_n in which the two numbers are chosen independently from the closed interval $[0,1]$ by means of the uniform distribution. By making n large our theorem, suitably normalized, suggests that the following strategies U^* and V^* are optimal in Ω. For player 1, cover a maximum number if and only if $x_1 + x_2 \geqslant 1$, and for player 2, accept the uncovered number if and only if it is at least $1/2$. Although we shall shortly prove that these strategies are indeed optimal the reader should not jump to the conclusion that they are obviously so. To see this we shall consider the following trivial game Δ_n. Each of two players simultaneously chooses a number from $\{1, 2, \ldots, n\}$. If they choose the same number no money changes hands. Otherwise, the player choosing the strictly larger number gets a dollar from his opponent unless he has chosen n, in which case he gives a dollar to his opponent. Clearly it is optimal for each player to choose the number $n - 1$. Now let Δ be the infinite game in which players 1 and 2 simultaneously choose numbers x and y, respectively, from the closed interval $[0,1]$. The payoff $M(x, y)$ to player 1 is given by

$$M(x,y) = \begin{cases} 0 & \text{for } x = y \\ -1 & \text{for } x < y < 1 \text{ or } y < x = 1 \\ 1 & \text{for } y < x < 1 \text{ or } x < y = 1. \end{cases}$$

By making n large in Δ_n and normalizing suitably, it might be thought that optimal strategies in Δ would be for each player to choose the number 1. However it is well known that Δ has no solution (see [**2**, p. 115]).

As promised, we now prove that the above strategies U^* and V^* are optimal in Ω by showing that each is a best reply to the other. If player 2 adopts V^* then, for a

best reply, player 1 covers a maximum if both numbers are at least $1/2$, a minimum if both are less than $1/2$, and either number otherwise. Hence it is easy to see that U^* is a best reply to V^*.

Now let player 1 adopt U^*. Suppose player 2 sees β, where $1 > \beta \geqslant 1/2$. This could have arisen via the pairs (x, β) and (β, x), where $\beta \leqslant x \leqslant 1$ or $0 \leqslant x < 1 - \beta$. Thus if player 2 rejects β his conditional expectation is

$$\left(\int_{\beta}^{1} x\,dx + \int_{0}^{1-\beta} x\,dx\right)\Big/(1 - \beta + 1 - \beta) = 1/2.$$

Hence he will not do worse by accepting β when $\beta \geqslant 1/2$.

Now suppose player 2 sees β where $0 < \beta < 1/2$. This could have arisen via the pairs (x, β) and (β, x) where $0 \leqslant x \leqslant \beta$ or $1 - \beta \leqslant x \leqslant 1$. Thus if player 2 rejects β his conditional expectation is

$$\left(\int_{0}^{\beta} x\,dx + \int_{1-\beta}^{1} x\,dx\right)\Big/(\beta + 1 - (1 - \beta)) = 1/2.$$

Hence he will not do worse by rejecting β when $\beta < 1/2$.

Thus V^* is a best reply to U^*.

We will find the value of Ω by calculating the expectation to player 2 when the players adopt the strategies U^* and V^*. Suppose player 1 sees x and y where $y \geqslant x \geqslant 1/2$. Then he covers y and player 2 accepts x. Thus the contribution to the expectation for these cases is

$$2\int_{1/2}^{1}\int_{1/2}^{y} x\,dx\,dy = 1/6.$$

Suppose player 1 sees x and y where $y \geqslant 1/2 > x$. Then, whichever he covers, player 2 gets y, and the contribution to expectation is

$$2\int_{1/2}^{1}\int_{0}^{1/2} y\,dx\,dy = 3/8.$$

Suppose player 1 sees x and y, where $1/2 > y > x$. Then he covers x and player 2 rejects y. Thus the contribution is

$$2\int_{0}^{1/2}\int_{0}^{y} x\,dx\,dy = 1/24.$$

Hence the value of Ω is $7/12$.

The reader may think that the corresponding games in which three numbers are chosen independently and player 1 covers two of them have equally simple solutions. Our investigations suggest that this is not the case and that a general solution would be difficult to find. Somewhat surprisingly, we have not been able to find any reference to such games in the literature and the nearest we have found is that to deception games, the study of which was initiated by Mark Thompson in an undergraduate thesis at Harvard University (see **[4]**). The only difference is that in deception games, instead of covering some of the numbers, player 1 tries to confuse his opponent by apparently changing their values. Little has been done on these problems, and they appear to be hard, but it has been proved recently that very general deception games always have a solution (see **[1]**).

REFERENCES

1. V. J. Baston and F. A. Bostock, Deception games, *Int. Journal of Game Theory* (submitted).
2. M. Dresher, Games of Strategy, Prentice Hall, London, 1961.
3. J. C. C. McKinsey, Introduction to the Theory of Games, McGraw-Hill, New York, 1952.
4. J. Spencer, A deception game, *Amer. Math. Monthly*, 80 (1973) 416-417.

A Note on Probabilities in Proofreading

FASHENG LIU
Computer Science Department, Shandong Teacher's University, P.R. China

Pólya [**1**] estimates the number M of misprints in a proofsheet using the moments methods for two independent proofreaders. It is sometimes desirable to estimate M for one proofreader proofreading twice, and it is also meaningful to estimate the operational level of proofreading (i.e., the probability that the proofreader notices any given misprint).

A proofreader reads the proofsheet two times by picking out misprints without replacement. Let M denote the number of misprints examined, p the probability that the proofreader notices any given misprint. Let X_1 and X_2 denote the numbers of misprints picked out in the first and second proofreadings. Using the moments methods as in [**1**], we have

$$Mp \sim X_1, \qquad (M - X_1)p \sim X_2$$

and so

$$M = \frac{X_1^2}{X_1 - X_2}, \qquad P = \frac{X_1 - X_2}{X_1}.$$

They are the desired estimates. Since the joint distribution of X_1 and X_2 is

$$P(X_1 = X_1, X_2 = X_2) = \frac{M!}{(M - X_1 - X_2)!X_1!X_2!} p^{X_1 + X_2}(1 - p)^{2M - 2X_1 - X_2},$$

the above estimators are also MLE (maximum likelihood estimator). The proofreader reads two proofsheets of the same book independently. As he finished, X_1 misprints were noticed in the first proofsheet, X_2 were noticed in 2nd proofsheet, X_{12} were noticed in both. So the desired estimates of M and p are:

$$M = \frac{X_1 X_2}{X_{12}} \qquad P = \frac{X_{12}(X_1 + X_2)}{2X_1 X_2}.$$

Two proofreaders read, independently of each other, the proofsheets of the same book. Let p_A and p_B be the operational levels of proofreading of the two proofreaders, A' and B'. A misprints were noticed by A', B misprints were noticed by B', C misprints by both. Then the desired estimates of p_A and p_B are:

$$P_A = \frac{C}{B}, \qquad P_B = \frac{C}{A}.$$

These are useful in quality control.

Reference

1. G. Pólya, Probabilities in proofreading, this MONTHLY, (1976) 42.

A Proof of the Lucas-Lehmer Test

MICHAEL I. ROSEN*
Department of Mathematics, Brown University, Providence, Rhode Island 02912

E. Lucas proposed and D. H. Lehmer proved a very elegant test for the primality of the Mersenne number $M_p = 2^p - 1$. Unfortunately, it is hard to find a proof in a modern text. Most books which state the result do not prove it, but do prove a related and weaker result, e.g. [**1**], [**2**], and [**4**]. In [**5**] a complete proof is given which follows the original proof of D. H. Lehmer (see [**3**]), but it is lengthy and detailed. In this note I give a fairly short proof using the simplest properties of algebraic numbers. Another proof using algebraic numbers was given long ago in [**6**].

Define a sequence S_n recursively by setting $S_1 = 4$ and $S_n = S_{n-1}^2 - 2$. So, $S_1 = 4$, $S_2 = 14$, $S_3 = 194$, etc.

THEOREM (LUCAS-LEHMER). *Let p be a prime number. Then, $M_p = 2^p - 1$ is a prime if and only if M_p divides S_{p-1}.*

Before giving the proof we need a few definitions. Let

$$\tau = \frac{1+\sqrt{3}}{\sqrt{2}}, \quad \bar{\tau} = \frac{1-\sqrt{3}}{\sqrt{2}}, \quad \omega = \tau^2 = 2+\sqrt{3}, \quad \text{and } \bar{\omega} = \bar{\tau}^2 = 2-\sqrt{3}.$$

Note that $\tau\bar{\tau} = -1$ and $\omega\bar{\omega} = 1$.

LEMMA. $S_m = \omega^{2^{m-1}} + \bar{\omega}^{2^{m-1}}$.

Proof. Let $T_m = \omega^{2^{m-1}} + \bar{\omega}^{2^{m-1}}$. Then $T_1 = \omega + \bar{\omega} = 4 = S_1$ and $T_{m+1} = T_m^2 - 2$. Thus, $T_m = S_m$ for all m.

LEMMA 2. *If we assume M_p is prime, then $\tau^{M_p+1} \equiv -1 \pmod{M_p}$.*

Proof. The congruence is understood to take place in the ring of algebraic integers. We temporarily set $q = M_p$.

Write $\sqrt{2}\,\tau = 1 + \sqrt{3}$, raise both sides to the qth power and take congruences modulo q. We find

$$\tau^q 2^{(q-1)/2}\sqrt{2} \equiv 1 + 3^{(q-1)/2}\sqrt{3} \pmod{q}.$$

Since $q \equiv -1 \pmod 8$ we have $2^{(q-1)/2} \equiv (2/q) \equiv 1 \pmod q$. Since $q \equiv 1 \pmod 3$ we have $3^{(q-1)/2} \equiv (3/q) \equiv -1 \pmod q$. It follows that

$$\tau^q \equiv \bar{\tau} \pmod{q} \text{ and } \tau^{q+1} \equiv \tau\bar{\tau} \equiv -1 \pmod{q}.$$

We can prove the theorem. We first prove that $S_{p-1} \equiv 0 \pmod{M_p}$ implies M_p is prime. This part of the proof makes no use of Lemma 2.

From the condition $S_{p-1} \equiv 0 \pmod{M_p}$ we find $\omega^{2^{p-2}} + \bar{\omega}^{2^{p-2}} \equiv 0 \pmod{M_p}$ so that $\omega^{2^{p-1}} \equiv -1 \pmod{M_p}$ and $\omega^{2^p} \equiv 1 \pmod{M_p}$.

Let l be a prime dividing M_p, and $\mathcal{O} = \mathbb{Z}[\sqrt{3}]$. The coset of ω in $(\mathcal{O}/l\mathcal{O})^*$ has order 2^p by the above congruences.

*This work was partically supported by NSF Grant DMS-8606407.

Suppose l splits in $\mathcal{O}$. Then it follows that

$$(\mathcal{O}/l\mathcal{O})^* \approx (\mathbb{Z}/l\mathbb{Z})^* \times (\mathbb{Z}/l\mathbb{Z})^*,$$

and so 2^p divides $l-1$. Thus, $l = 1 + 2^p k$ for some $k \geqslant 1$. This is impossible because it implies $l \geqslant 1 + 2^p > M_p$.

Suppose l stays prime in $\mathcal{O}$. Then $(\mathcal{O}/l\mathcal{O})^*$ has order $l^2 - 1$ and 2^p divides $l^2 - 1 = (l-1)(l+1)$. If $l \equiv 1 \pmod 4$ we must have 2^{p-1} divides $l-1$, or $l = 1 + 2^{p-1}k$ for some $k \geqslant 1$. This cannot happen since $2l \geqslant 2 + 2^p > M_p$. If $l \equiv 3 \pmod 4$, then 2^{p-1} divides $l+1$ so that $l = -1 + 2^{p-1}k$ for some $k \geqslant 1$. One cannot have $k = 1$ since $2^{p-1} - 1$ does not divide $2^p - 1$. The only possible case is $k = 2$. Then, $l = 2^p - 1 = M_p$, and so M_p is prime.

Finally, suppose M_p is prime. From Lemma 2 we have

$$\tau^{M_p+1} \equiv -1 \left(\operatorname{mod} M_p\right) \quad \text{or } \tau^{2^p} + 1 \equiv 0 \left(\operatorname{mod} M_p\right).$$

Since $\tau^2 = \omega$, $\omega^{2^{p-1}} + 1 \equiv 0 \pmod{M_p}$. Multiply both sides of this congruence with $\bar{\omega}^{2^{p-2}}$ and recall $\omega\bar{\omega} = 1$. The result is

$$S_{p-1} = \omega^{2^{p-2}} + \bar{\omega}^{2^{p-2}} \equiv 0 \left(\operatorname{mod} M_p\right).$$

REFERENCES

1. G. H. Hardy and E. M. Wright, The Theory of Numbers, Oxford at the Clarendon Press, 1965.
2. L. K. Hua, Introduction to the Theory of Numbers, Springer-Verlag, Berlin-Heidelberg-New York, 1982.
3. D. H. Lehmer, On Lucas' test for the primality of Mersenne's numbers, *J. of the London Math. Soc.*, 10 (1935) 162–165.
4. H. Riessel, Prime Numbers and Computer Methods for Factorization, Birkhauser, Boston-Basel-Stuttgart, 1985.
5. W. Sierpinski, Elementary Theory of Numbers, Polski Akademic Nauk., Warsaw, 1964.
6. A. E. Western, On Lucas' and Pepin's tests for the primeness of Mersenne numbers, *J. of the London Math. Soc.*, 7 (1932) 130–137.

The Behaviour of Means Under Equal Increments of Their Variables

J. ACZÉL

Department of Pure Mathematics, University of Waterloo, Waterloo, Ont. Canada N2L 3G1

ZS. PÁLES

Institute of Mathematics, L. Kossuth University, H-4010 Debrecen, Hungary

In [**6**], after noting that the arithmetic ($M = A$), geometric ($M = G$), harmonic ($M = H$) means, and the root-mean square (or quadratic mean $M = Q$, denoted in [**6**] by R) all satisfy

$$M(a_1 x, \dots, a_n x) = M(a_1, \dots, a_n)x \tag{1}$$

(here and in what follows, except in Proposition 4, $a_1, \dots, a_n, x$ are supposed to be positive), and

$$A(a_1 + x, \dots, a_n + x) = A(a_1, \dots, a_n) + x, \tag{2}$$

the authors prove separately the following inequalities:

$$\left.\begin{aligned} G(a_1+x,\dots,a_n+x) &> G(a_1,\dots,a_n)+x \textit{ and}\\ G(a_1+x,\dots,a_n+x) &- x \textit{ increases with } x, \end{aligned}\right\} \tag{3}$$

$$\left.\begin{aligned} H(a_1+x,\dots,a_n+x) &> H(a_1,\dots,a_n)+x \textit{ and}\\ H(a_1+x,\dots,a_n+x) &- x \textit{ increases with } x, \end{aligned}\right\} \tag{4}$$

$$\left.\begin{aligned} Q(a_1+x,\dots,a_n+x) &< Q(a_1,\dots,a_n)+x \textit{ and}\\ Q(a_1+x,\dots,a_n+x) &- x \textit{ decreases with } x. \end{aligned}\right\} \tag{5}$$

The authors' next aim is to dispel the worry that $Q(a_1+x,\dots,a_n+x) - A(a_1+x,\dots,a_n+x)$, which in view of (5) and (2) decreases, could go below 0 in contradiction to $Q > A$ and similar worries in opposite direction concerning (4) and (3). For this purpose the authors prove

$$\lim_{x\to\infty}\left[M(a_1+x,\dots,a_n+x) - A(a_1+x,\dots,a_n+x)\right] = 0$$

or, what is equivalent by (2),

$$\lim_{x\to\infty}\left[M(a_1+x,\dots,a_n+x) - x\right] = A(a_1,\dots,a_n),$$

again individually for $M = G$, $M = H$, and $M = Q$ (for $M = A$ this is obvious from (2)).

In what follows we give short elementary proofs for *all root-mean powers*

$$\left.\begin{aligned} &M = M_p(a_1,\dots,a_n) = \left(\sum_{i=1}^n a_i^p\Big/n\right)^{1/p} && (p\neq 0)\\ \text{and}\\ &M = M_0(a_1,\dots,a_n) = G(a_1,\dots,a_n) = \prod_{i=1}^n a_i^{1/n} && (p=0) \end{aligned}\right\} \tag{6}$$

(clearly $M_1 = A$, $M_{-1} = H$, $M_2 = Q$) of the following statements (and, later, some *generalizations*).

PROPOSITION 1. *Equation* (1) *holds for all* M_p $(p\in\mathbb{R})$, *as defined in* (6).

PROPOSITION 2. *As a function of* x,

$$M_p(a_1+x,\dots,a_n+x) - x$$

decreases or increases (*or is constant*) *and is smaller or greater than* (*or equal to*) $M_p(a_1,\dots,a_n)$ *according as* $p \gtrless 1$ (*or* $p = 1$). *This increase or decrease and the inequalities are strict unless* $a_1 = \cdots = a_n$.

PROPOSITION 3. *For all* $p\in\mathbb{R}$

$$\lim_{x\to\infty}\left[M_p(a_1+x,\dots,a_n+x) - x\right] = A(a_1,\dots,a_n) = \frac{a_1+\cdots+a_n}{n}. \tag{7}$$

Proofs.

PROP. 1: Trivial. □

Actually (see, e.g., [**5**]) the equation (1) *characterizes the root-mean powers* (6) *among the quasi-arithmetic means* (*cf.* [**1**])

$$M(x_1,\dots,x_n)=\phi^{-1}\left(\sum_{i=1}^{n}\phi(x_i)\Big/n\right)$$

(where ϕ is continuous and strictly monotonic).

PROP. 2 (All sums and later products are with respect to i and go from 1 to n):

$$\left.\begin{aligned}&\frac{d}{dx}\left[M_p(a_1+x,\dots,a_n+x)-x\right]=\frac{d}{dx}\left[\frac{\left(\Sigma(a_i+x)^p\right)^{1/p}}{n^{1/p}}-x\right]\\&=\frac{1}{n^{1/p}}\left(\sum(a_i+x)^p\right)^{(1/p)-1}\sum(a_i+x)^{p-1}-1\gtreqless 0\ \textit{if}\ p\lesseqgtr 1.\end{aligned}\right\}\tag{8}$$

Indeed, for $p>1$, we have with the notation $y_i=(a_i+x)^p$ $(i=1,\dots,n)$, by Hölder's inequality (see, e.g., [**5**])

$$\sum 1^{1/p}y_i^{1-(1/p)}\leqslant\left(\sum 1\right)^{1/p}\left(\sum y_i\right)^{1-(1/p)}$$

with equality iff $y_1/1=\cdots=y_n/1$, i.e., $a_1=\cdots=a_n$. For $p<1$ the inequality reverses. Accordingly,

$$M_p(a_1+x,\dots,a_n+x)-x\lessgtr M_p(a_1,\dots,a_n)\ \text{if}\ p\gtrless 1.\quad\square$$

An even simpler proof is the following. By Minkowski's inequality (see, e.g., [**5**]) we have for $p>1$, whenever $x<y$,

$$\begin{aligned}&M_p(a_1+x,\dots,a_n+x)-x=M_p(a_1+x,\dots,a_n+x)\\&+M_p(y-x,\dots,y-x)-y\geqslant M_p(a_1+y,\dots,a_n+y)-y\end{aligned}$$

with equality iff $a_1=\cdots=a_n$. If $p<1$, the inequality signs are reversed. Minkowski's inequality gives also directly $M_p(a_1+x,\dots,a_n+x)\lessgtr M_p(a_1,\dots,a_n)+x$ for $p\gtrless 1$, that is (in generalization of (3), (4) and (5)),

$$\left.\begin{aligned}M_p(a_1+x,\dots,a_n+x)-x\lessgtr M_p(a_1,\dots,a_n)\\ \text{for } p\gtrless 1,\ \text{except } \textit{if}\ a_1=\cdots=a_n.\end{aligned}\right\}\ \square\tag{9}$$

PROP. 3: Using the notation $t=1/x$ and the Bernoulli-l'Hospital theorem,

$$\begin{aligned}\lim_{x\to\infty}\left[\left(\frac{\Sigma(a_i+x)^p}{n}\right)^{1/p}-x\right]&=\lim_{t\to0}\frac{\left(\Sigma(a_it+1)^p\right)^{1/p}-n^{1/p}}{n^{1/p}t}\\&=\lim_{t\to0}\frac{1/p\left(\Sigma(a_it+1)^p\right)^{(1/p)-1}p\Sigma(a_it+1)^{p-1}a_i}{n^{1/p}}\\&=\frac{n^{(1/p)-1}\Sigma a_i}{n^{1/p}}=\frac{\Sigma a_i}{n}.\quad\square\end{aligned}$$

Remarks. (i) The second proof of Proposition 2 remains unchanged with $p = 0$. Since (see, e.g., [5]) $\lim_{p \to 0} M_p = M_0$, Proposition 3 follows also for $p = 0$ and the first proof of Proposition 2 'almost' covers the case $p = 0$: except the uniqueness of the equality case. However, a direct proof for $M_0 = G$ is very simple too:

$$\frac{d}{dx}[G(a_1 + x, \dots, a_n + x) - x] = \frac{d}{dx}\left(\prod (a_i + x)^{1/n} - x\right)$$
$$= \frac{1}{n}\prod (a_i + x)^{1/n} \sum \frac{1}{a_i + x} - 1 > 0,$$

because the geometric mean $\prod(a_i + x)^{1/n}$ is greater than the harmonic mean $n/\sum(a_i + x)^{-1}$ if not all a_i are equal.

(ii) All three propositions remain true (with unchanged or slightly modified proofs) for *weighted* root-mean powers $((\sum q_i a_i^p)^{1/p}$ for $p \neq 0$, $\prod a_i^{q_i}$ for $p = 0$ $(q_i > 0;\ i = 1, \dots, n;\ \sum q_i = 1))$, with $A(a_1, \dots, a_n) = \sum q_i a_i$ in (7).

(iii) In generalization of Proposition 3 and with a very similar proof we have the following.

PROPOSITION 4. *Let I be one of the intervals $\mathbb{R}_+ =]0, \infty[$, $\mathbb{R} =]-\infty, \infty[$, and let M be homogeneous, that is,*

$$\left.\begin{array}{c} M(v_1 x, \dots, v_n x) = M(v_1, \dots, v_n)x \\ \textit{for all } v_i \in I\ (i = 1, \dots, n) \textit{ and for all positive } x, \end{array}\right\} \tag{10}$$

let M satisfy

$$M(1, \dots, 1) = 1 \tag{11}$$

and be differentiable at $(1, \dots, 1)$. Write $q_i = \dfrac{\partial M}{\partial v_i}(1, \dots, 1)(i = 1, \dots, n)$. Then

$$\lim_{x \to \infty} [M(a_1 + x, \dots, a_n + x) - x] = \sum q_i a_i \textit{ and } \sum q_i = 1. \tag{12}$$

Proof. Introducing again $t = 1/x$ we get from (10) (by using the definition of the derivative instead of the Bernoulli-l'Hospital theorem)

$$\lim_{x \to \infty} [M(a_1 + x, \dots, a_n + x) - x] = \lim_{x \to \infty} x\left[M\left(\frac{a_1}{x} + 1, \dots, \frac{a_n}{x} + 1\right) - 1\right]$$
$$= \lim_{t \to 0} \frac{M(a_1 t + 1, \dots, a_n t + 1) - 1}{t}$$
$$= \sum \frac{\partial M}{\partial v_i}(1, \dots, 1) a_i = \sum q_i a_i.$$

From (10) and (11) we get $M(x, \dots, x) = x$ for all $x > 0$ and thus

$$\sum \frac{\partial M}{\partial v_i}(x, \dots, x) = 1,$$

in particular $\sum q_i = 1$. □

Note. Here $a_1, \dots, a_n$ do not need to be positive anymore; neither do all the q_i. However, if M is symmetric, then $q_i = \cdots = q_n = 1/n$ and the limit is $A = \sum a_i/n$ as before.

(iv) Losonczi [7] gave necessary and sufficient conditions for quasi-linear means with weight functions

$$M(a_1,\ldots,a_n) = \phi^{-1}\left(\sum f(x_i)\phi(x_i)/\sum f(x_i)\right)$$

to satisfy (cf. (9))

$$M(a_1+u,\ldots,a_n+u) - u > M(a_1,\ldots,a_n) \quad \text{for } u > 0. \tag{13}$$

By the way, *it follows from (13) that the function $u \mapsto M(a_1+u,\ldots,a_n+u) - u$ is strictly increasing*. Indeed, take an arbitrary $y > x$ and $u = y - x > 0$. Then (13) yields

$$\begin{aligned} M(a_1+y,\ldots,a_n+y) - y &= M(a_1+x+(y-x),\ldots,a_n+x+(y-x)) \\ &\quad -(y-x) - x \\ &> M(a_1+x,\ldots,a_n+x) - x, \end{aligned}$$

as asserted.

(v) In the second proof of Proposition 2, Minkowski's inequality is used. In [8] Minkowski type inequalities were proved under certain conditions for mean values even more general than quasi-linear means with weight functions. Proposition 2 holds in all those cases.

We think that the above, in particular Propositions 1, 2, 3, 4, Remark (ii) and the second part of (iv), show that more general statements may have at least as simple elementary proofs as their particular cases.

(vi) Independently, Brenner and Carlson [3] have proved (personal communication) a result somewhat stronger than (12), namely,

$$x\left[M(a_1+x,\ldots,a_n+x) - x - \sum q_i a_i\right] \text{ is bounded as } x \to \infty$$

under somewhat stronger conditions than those in Proposition 4 above, supposing, in particular, instead of differentiability at $(1,\ldots,1)$, the existence and continuity of all second order partial derivatives (everywhere). Previously Brenner [2] has proved Proposition 3 for positive integer p by binomial and other expansions and for noninteger p by the monotonicity of $p \mapsto M_p(a_1,\ldots,a_n)$. Recently Bullen [4] has proved (personal communication) Proposition 2 independently by a variant of our first proof, also using that M_p is increasing in p, and Proposition 3, again by use of binomial and multinomial expansions (both for weighted means; cf. the above Remark (ii)).

The simplicity of the proof of Proposition 4 and of the second proof of Proposition 2 here may be of some comparative interest.

REFERENCES

1. J. Aczél, Lectures on Functional Equations and Their Applications. Academic Press, New York and London, 1966.
2. J. L. Brenner, Limits of means for large values of the variable, *Pi Mu Epsilon J.*, 8 (1985) 160–163.
3. J. L. Brenner and C. Carlson, Homogeneous mean values: weights and asymptotics, *J. Math. Anal. Appl.*, 123 (1987) 265–280.
4. P. S. Bullen, Averages still on the move, *Math. Mag.*, to appear.
5. G. H. Hardy, J. E. Littlewood, and G. Pólya, Inequalities, Cambridge Univ. Press, 1952.
6. L. Hoehn and I. Niven, Averages on the move, *Math. Mag.*, 58 (1985) 151–156.
7. L. Losonczi, Subhomogene Mittelwerte, *Acta Math. Acad. Sci. Hung.*, 22 (1971) 187–195.
8. Zs. Páles, General inequalities for quasideviation means, *Aequationes Math.*, to appear.

EDITED BY JOAN P. HUTCHINSON AND STAN WAGON

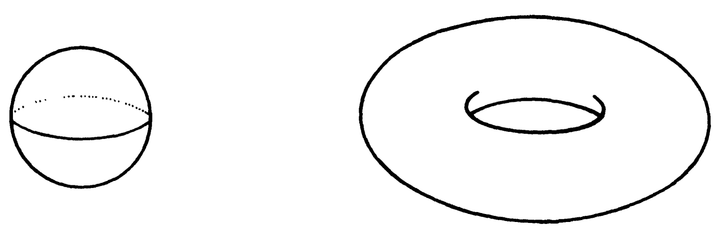
The Classification of Surfaces

PETER ANDREWS
Department of Mathematics, Wilfrid Laurier University, Waterloo, Ontario, Canada N2L 3C5

1. Introduction. The classification of surfaces is one of the truly beautiful theorems of modern mathematics. There are many proofs in the literature already, most of which are quite elementary. This note presents yet another proof, one that involves some very simple graph theory, an area increasingly popular with, and familiar to, students today. It also involves rudimentary surgery and thus could serve as an introduction to some of the methods of modern differential topology.

A surface is a compact two-manifold without boundary. For those readers for whom this is not very helpful, [**7**, Chap. 1] and [**4**, Chap. 1, 2] both give a very leisurely, well-illustrated explanation.

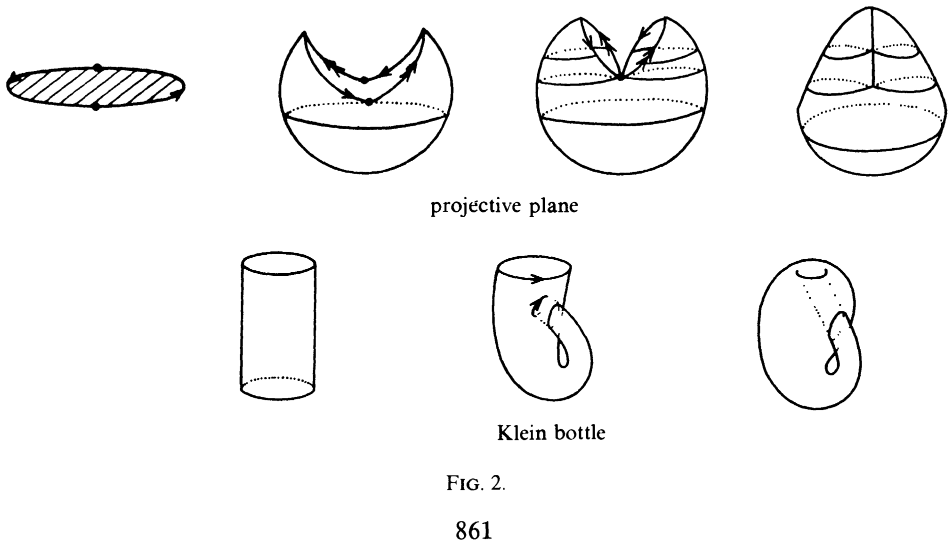
Perhaps the most familiar surfaces are the sphere and the torus (FIG. 1).



sphere torus

FIG. 1.

Two more exotic surfaces are the projective plane and the Klein bottle (FIG. 2; see also [**6**, pp. 64–67]).



projective plane

Klein bottle

FIG. 2.

Most mathematicians are familiar with the Möbius strip (FIG. 3).

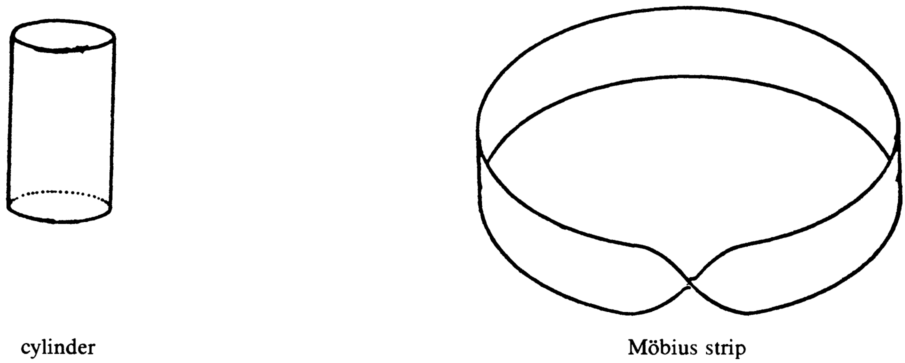
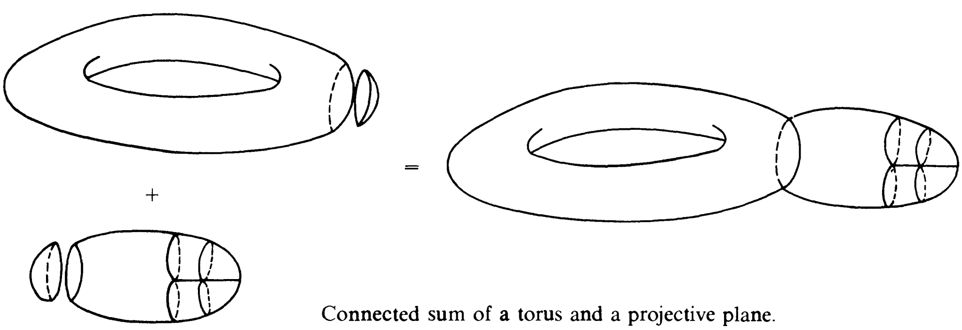


FIG. 3.

This is not a surface but, like the cylinder, a surface-with-boundary. The boundary of the cylinder is, of course, two disjoint circles, while the boundary of the Möbius strip is just a single circle. The Möbius strip will be quite important in our proof because it is so closely related to the projective plane. In fact, if the interior of a small disc is removed from a projective plane the resulting surface-with-boundary is a Möbius strip (see [**5**, p. 33]).

Any two surfaces can be combined to produce a third by forming what is called their connected sum (see [**5**, p. 79]). A small open disc is removed from each surface, and the two resulting boundary circles are then glued together (FIG. 4).



Connected sum of a torus and a projective plane.

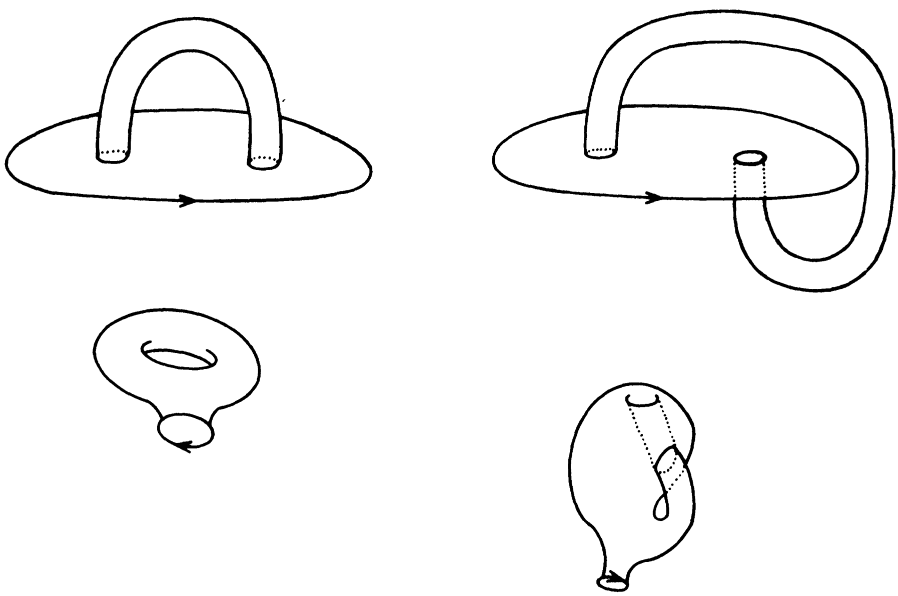
FIG. 4.

The simplest and weakest form of the classification theorem says that any surface can be built from the few examples described above using only the operation of the connected sum.

THEOREM. *Any surface is homeomorphic to the connected sum of a sphere, m tori, n projective planes, and r Klein bottles, where m, n, and r are greater than or equal to zero.*

The basic idea of this proof is the same as that of [**5**, pp. 88–89]. We attempt to simplify the surface by removing tori, projective planes, or Klein bottles. Begin by

looking for a nonseparating simple closed curve γ on Σ, that is, a curve γ for which $\Sigma - \gamma$ is still connected. If W is a small open neighborhood of γ on Σ, then $\overline{W}$ (the closure of W) is either a cylinder or a Möbius strip. Consider $\Sigma - W$ the surface-with-boundary obtained by removing W from Σ. The boundary of $\Sigma - W$ will be one or two circles since $\mathrm{bdry}(\Sigma - W) = \mathrm{bdry}(\overline{W})$. Glue a disc (or discs) onto the boundary circle(s) of $\Sigma - W$ to get a new surface Σ'. In essence we go from Σ to Σ' by removing a cylinder (handle) or a Möbius strip. A little care must be taken here. The handle could have been attached nicely (FIG. 5(a)) or badly (FIG. 5(b)). It is a good exercise to show that the nice handle in FIG. 5(a) is homeomorphic to a torus with a disc removed and that the bad handle in FIG. 5(b) is homeomorphic to a Klein bottle with a disc removed.



(a) "nice" handle (b) "bad" handle

FIG. 5.

We continue by looking for a nonseparating simple closed curve γ' on Σ' and repeating the process. This leaves two major questions: Will we ever reach a surface with no nonseparating curves? What can we say about such a surface? The answers are quite simple. After a finite number of these surgeries, we *will* obtain a surface with no nonseparating closed curves and such a surface is homeomorphic to a sphere. The verification of these statements is usually rather lengthy, though.

In this proof we use triangulated surfaces and some graph theory to describe a very concrete execution of the program just sketched. We show where to look for the non-separating curve γ and demonstrate easily that the procedure produces a sphere in a finite number of steps. For convenience, we use a slightly different technique for the surgery. Instead of removing an entire neighborhood of γ, we merely cut the surface open along γ. The surface-with-boundary created is homeomorphic to the one described above and so the ultimate effect is the same.

The idea of this proof came from Paul Melvin, who credits it originally to E. C. Zeeman. (The referee has pointed out that a similar version of this proof [also credited to Zeeman] appears in an excellent textbook by M. A. Armstrong [**1**]. Regrettably, this book does not seem to be very well known or readily available in North America. We hope that those readers who find this article useful will be able to examine Armstrong's text in its entirety. The author would also like to acknowledge financial support from Williams College, McMaster University, and Wilfrid Laurier University.)

2. Some Graph Theory. For purposes of this note, we assume that *all graphs are connected*. Recall that a *spanning tree* in a graph G is a subgraph of G that is a tree containing all the vertices of G. It is easy to see that every graph has a spanning tree.

Let $V(G)$ (resp., $E(G)$) denote the number of vertices (resp., edges) in a graph of G.

DEFINITION. The *cyclomatic number* of a graph G, denoted by $C(G)$, is defined by the equation

$$C(G) = 1 - V(G) + E(G).$$

The cyclomatic number measures the number of independent cycles in G. We need only one of its properties. Some deeper results about $C(G)$ may be found in [**2**, Chap. 2].

PROPOSITION. $C(G) \geqslant 0$. *Furthermore*, $C(G) = 0$ *if and only if* G *is a tree.*

Proof. The proof follows quite easily using induction on the number of edges.

3. Proof of the Theorem. As in most proofs of this theorem, we use combinatorial techniques and so assume that all surfaces are triangulated. It is a nontrival result that this introduces no loss of generality in dimension two (see, e.g., [**3**, §16.a]).

DEFINITION. A *surface* is a finite collection of triangles in some Euclidean space such that:

(i) the intersection of any two triangles is empty, a single vertex, or a single edge;
(ii) any edge is the intersection of exactly two triangles; and
(iii) the triangles incident with a given vertex can be ordered $\Delta_1, \Delta_2, \dots, \Delta_k$ so that Δ_i has exactly one edge in common with Δ_{i+1}, for $1 \leqslant i < k$, and Δ_k has exactly one edge in common with Δ_1.

For a surface Σ, let $\Delta_1, \Delta_2, \dots, \Delta_n$ be a list of all the triangles of Σ. The collection of vertices and edges of these triangles forms a graph that we also denote by Σ. Let B_i be the barycenter (centroid) of Δ_i. Form a new graph G_Σ whose vertices are the barycenters $B_1, B_2, \dots, B_n$. There will be an edge B_iB_j in G_Σ if and only if Δ_i and Δ_j have an edge e in common. In this case we say that B_iB_j crosses e.

Let T be any spanning tree of G_Σ and K_T be the subgraph of Σ consisting of all the vertices of Σ and those edges of Σ not crossed by T. It is easy to verify that K_T is connected since T is a tree.

It is perhaps helpful to think of T as lying along the edges of $b(\Sigma)$, the first barycentric subdivision of Σ. It is not, however, a subgraph of $b(\Sigma)$ because the

vertices of T are just the barycenters of the triangles in Σ. We can use $b^2(\Sigma)$, the second barycentric subdivision of Σ, to produce neighborhoods $U \supset T$ and $V \supset K_T$ on the surface Σ satisfying:

(i) $U \cap V = \phi$,
(ii) $\overline{U} \cup \overline{V} = \Sigma$, and
(iii) bdry $U =$ bdry V.

For example, U will be T together with the interiors of all the triangles and edges of $b^2(\Sigma)$ that intersect T. There is a similar description for V. We should remark here that $\overline{U}$ is homeomorphic to a two-dimensional disc since T is a tree.

FIGURE 6 shows a triangulation of the projective plane. (Note that there are identifications on the boundary of the hexagon.) One of the possible choices of T is drawn with heavy lines. The corresponding K_T is drawn with broken lines. The shaded region is U and the unshaded one is V.

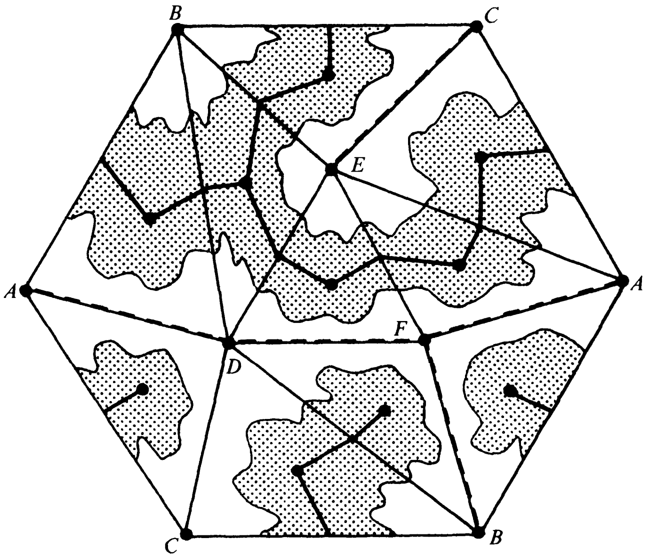


FIG. 6.

Now focus on the cyclomatic number $L = C(K_T)$. We consider two cases.

Case 1: $L = 0$. In this case K_T is a tree. Then $\overline{V}$ is a disc, as observed above. Thus Σ is the union of two discs, $\overline{U}$ and $\overline{V}$, joined by an identification of their common boundary. This means that Σ is homeomorphic to a sphere.

Case 2: $L \neq 0$. In this case, K_T is not a tree so there must be a cycle γ in K_T. This is the curve on which we do surgery.

LEMMA. *The cycle γ does not separate Σ.*

Proof. The tree T has a vertex in each triangle of Σ and does not cross γ. Thus any two points of $\Sigma - \gamma$ can be joined by a path in $\Sigma - \gamma$.

Instead of removing a whole neighborhood of γ, we just cut the surface open along γ to create a triangulated surface-with-boundary Σ_γ. Just as in our earlier

discussion, the boundary of Σ_γ consists of one circle if γ has Möbius strip neighborhood and of two circles if γ has a cylinder neighborhood.

In the example of FIG. 6, $L = 1$ and essentially the only choice of γ is the cycle ADFA. The neighborhood of γ is shown in FIG. 7. It is obviously a Möbius strip.

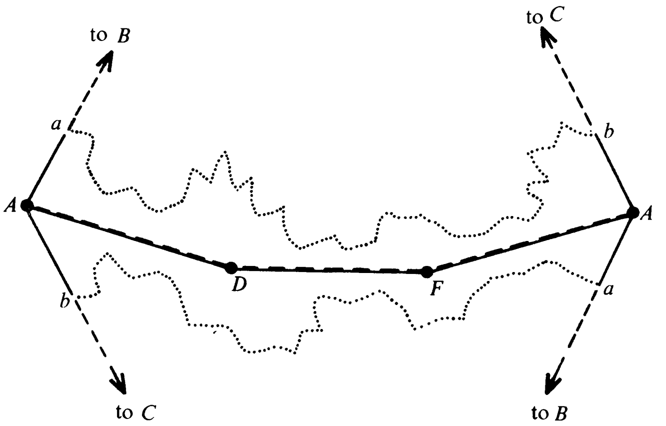


FIG. 7.

The next step is to cap off the boundary of Σ_γ with a suitably triangulated disc (or discs) as shown in FIG. 8. This creates a new surface Σ'. The heavy lines in FIGURE 8 show how the existing tree T can be extended into each new triangle to get a new tree T' that will be a spanning tree for $G_{\Sigma'}$. We then compute $L' = C(K_{T'})$ by simply counting the vertices and edges added to or lost from K_T to produce $K_{T'}$.

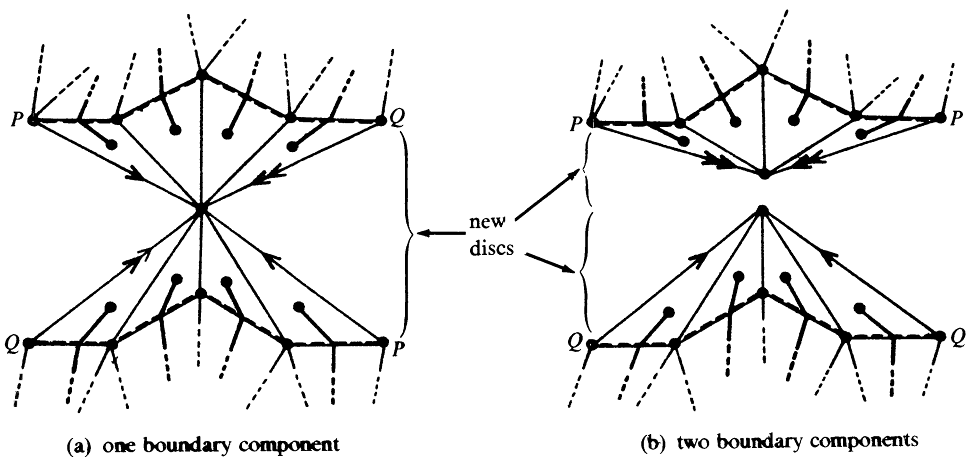


FIG. 8.

LEMMA. *If Σ_γ has one boundary circle, then $L' = L - 1$; if Σ_γ has two boundary circles, then $L' = L - 2$.*

The whole procedure can now be iterated, decreasing the value of L each time. In a finite number of steps, then, we arrive at the $L = 0$ case, which corresponds to a sphere. As outlined in Section 1, each surgery (the process of cutting along γ and capping off the boundary of Σ_γ) is equivalent to removing a torus, a projective plane or a Klein bottle. This completes the proof of the theorem.

REFERENCES

1. M. A. Armstrong, Basic Topology, McGraw-Hill (UK), London, 1979.
2. C. Berge, Graphs and Hypergraphs, North Holland, Amsterdam, 1976.
3. C. O. Christenson and W. L. Voxman, Aspects of Topology, Marcel Dekker, New York, 1977.
4. H. B. Griffiths, Surfaces, Cambridge University Press, Cambridge, 1976.
5. C. Kosniowski, A First Course in Algebraic Topology, Cambridge University Press, Cambridge, 1980.
6. J. Stillwell, Classical Topology and Combinatorial Group Theory, Springer-Verlag, New York, 1980.
7. J. R. Weeks, The Shape of Space, Marcel Dekker, New York, 1985.

Another Required Reading Program for Mathematics Majors

JAMES C. REBER
Department of Mathematics, Indiana University of Pennsylvania, Indiana, PA 15705

I appreciated the article by Robert L. Brabenec, "A Required Reading Program for Mathematics Majors" [this MONTHLY, April 1987]. For precisely the reasons he presented we have introduced a reading program at Indiana University of Pennsylvania. The differences between programs, and respective advantages and disadvantages might be worth noting.

Our program is tied to individual courses required for all majors, and uses readings on reserve in the library. Each instructor teaching a course in the program requires the readings for that course, specifies how the requirement is met (typically with a short reaction paper), and determines how the assignment will be graded. General readings are used in courses populated primarily with mathematics majors; readings explicitly tied to the course are used in courses that include many non-mathematics majors.

One disadvantage of this approach: though encouraged to complete the assignment early in the semester, students often complete it amid the pressure of other work.

Advantages: students minoring in mathematics or taking just a few courses complete some readings; faculty are more likely to integrate the readings into the course; there is no departmentwide bookkeeping; and the program is easy to start since it can be phased in several courses at a time.

The explicit readings we use might be instructive, though I agree with Brabenec's comment, "The specific content of what our students are reading is not nearly as important as the fact that they are reading."

Calculus I. Read two of three essays on the nature of mathematics from *Mathematics: People/Problems/Results*: "Math and Creativity" by Alfred Adler; "The Meaning of Math" by Morris Kline; "Math as Creative Art" by Paul Halmos.

Calculus II. Read specified selections from either *The Mathematical Experience* by Davis and Hersh or from *How to Solve It* by Pólya.

Calculus III. Read specified selections from either *Infinity* by Lieber or from *Bridges to Infinity* by Guillen.

Introduction to Linear Algebra. Read three articles from *Scientific American* as follows. Either "Linear Programming," August 1954, or "The Allocation of Resources by Linear Programming," June 1981. One of these articles: "Input-Output Economics", October 1951; "The Structure of U.S. Economy," April 1965; or "The World Economy by the Year 2000," September 1980. One article from the "Mathematics and Modern World" issue of September 1964. ("Math in Social Sciences" is most relevant to this course.)

Introduction to Algebraic Structures. Read either *Flatland* by Abbott, or specified selections from Kline's *Mathematics, The Loss of Certainty* and from Kasner and Newman's *Mathematics and the Imagination.*

A Principal Ideal Domain That Is Not a Euclidean Domain

OSCAR A. CÁMPOLI
Facultad de Matemática, Astronomía y Física, Valparaíso y R. Martínez Ciudad Universitaria, 5000 Córdoba, Argentina

Introduction. In most advanced undergraduate and graduate algebra texts a very simple argument is used to show that a Euclidean domain is a principal ideal domain (PID). And then it is mentioned that the converse is not true, sometimes together with the claim that the subring $A = \mathbb{Z}[\theta] = \{a + b\theta | a, b \in \mathbb{Z},\ \theta = (1 + \sqrt{-19})/2\}$ of the complex numbers is a PID but is not a Euclidean domain. I have not been able to find a proof, accessible to beginning graduate students, in any standard reference (e.g., [**1, 2, 3, 4**]).

In what follows it is shown in an elementary fashion that A has both properties.

The proof that A is not a Euclidean domain is in [**5**] but we use here a shorter argument suggested by the referee.

One way to see that A is a PID can be found in algebraic number theory books where the class number of the field $\mathbb{Q}(\sqrt{-19})$ is computed. The proof given here uses that A is "almost" a Euclidean domain in the sense that it has a "generalized" Euclidean algorithm. A criterion (sometimes attributed to Dedekind and Hasse) is then proven and used to show that A is a PID.

A is not Euclidean. In general, it is not clearly stated what Euclidean domains are. A definition is as follows:

A *Euclidean domain* consists of an integral domain A together with a map $|\ |: A \to \mathbb{Z}$ (the *Euclidean norm*) that satisfies the following conditions:

(i) $|\ |(a) = |a| \geqslant 0$ for all $a \in A$; $|a| = 0$ if and only if $a = 0$.

(ii) $|ab| = |a| \cdot |b|$ for all $a, b \in A$.

(iii) (Euclidean algorithm) Given $a, b \in A$, $b \neq 0$, there exist $q, r \in A$ so that $a = qb + r$ with $|r| < |b|$.

It is interesting to note that condition (ii) of the definition can be weakened to

(ii′) $|a| \leqslant |b|$ whenever a divides b (for nonzero b),

which follows easily from (ii). In fact (ii′) will be used instead of (ii).

To show that A is not Euclidean it is sufficient to prove that A does not admit a function $|\,|$ satisfying the three stated properties. Thus assume that $|\,|$ is a Euclidean norm in A. This leads to a contradiction.

Indeed, let U be the set of nonzero elements in A with minimal norm. Since every unit of A divides every nonzero element, (ii′) implies that every unit is in U and (iii) implies that every element of U divides every nonzero element of A; so U consists precisely of the units of A.

We next show that $U = \{1, -1\}$. In order to prove this and other assertions a few specific calculations in the ring A are needed.

The following identities can be proved directly from the definition of $\theta = (1 + \sqrt{-19})/2$. For $a \in A$, $\bar{a}$ denotes the complex conjugate of the complex number a.

(I) $\bar{\theta} = 1 - \theta$

(II) $\theta\bar{\theta} = 5$

(III) $\theta^2 = \theta - 5$

(IV) For any $x = a + b\theta \in A$, $\theta x = -5b + (a + b)\theta$.

From (I) it follows that A is closed under complex conjugation. Identity (II) implies that the integer 5 is not a prime in A. Later it will be clear that θ is not a unit in A and it will then follow that 5 is reducible in A. From (III) it follows that $\theta^2 \in A$ and hence A is closed under complex multiplication (a fact not obvious from the definition of A).

If $N(z) = z\bar{z}$ is the usual complex norm, then the preceding identities yield:

(V) $N(a + b\theta) = (a + b\theta)\overline{(a + b\theta)} = a^2 + ab + 5b^2$.

Moreover, the function $N: A \to \mathbb{Z}$ satisfies

(a) $N(xy) = N(x)N(y)$ for all $x, y \in A$, and

(b) $N(x) \geqslant 0$ for all $x \in A$ and $N(x) = 0$ if and only if $x = 0$.

This immediately implies that if an element $a + b\theta \in A$ is a unit then $a^2 + ab + 5b^2 = N(a + b\theta) = 1$ and hence, if $ab \geqslant 0$, then $b = 0$ and $a = \pm 1$. Also, since $a + b\bar{\theta} = a + b - b\theta$ and $1 = N(a + b\theta) = N(a + b\bar{\theta}) = (a + b)^2 - ab + 4b^2$, it follows that when $ab \leqslant 0$ then again $b = 0$ and $a = \pm 1$. This concludes the proof of the fact that $U = \{1, -1\}$.

Now assume that m is of minimal norm among the elements of A different from $0, 1, -1$. Condition (iii) implies that $2 = qm + r$, with $|r| < |m|$; therefore r is one of $0, 1$, or -1. Hence either m divides 2 or m divides 3. We claim that m must then be one of ± 2, ± 3.

This claim is a consequence of the fact that 2 and 3 are primes in A, which is shown as follows. Suppose $2 = (a + b\theta)(c + d\theta)$. Then $4 = N(2) = N(a + b\theta) N(c + d\theta)$ and assuming that $a + b\theta$, $c + d\theta$ are not units in A, it follows that

$$2 = N(a + b\theta) = a^2 + ab + 5b^2 = N(a + b\bar{\theta}) = (a + b)^2 - ab + 4b^2.$$

Therefore, considering the cases $ab \geqslant 0$ and $ab < 0$, we conclude that b and d each equal zero.

Thus $2 = (a + b\theta)(c + d\theta) = ac$ is an integral factorization. Since 2 is a prime in $\mathbb{Z}$, 2 is a prime in A. A similar argument shows that 3 is also a prime in A.

Now, again using (iii), θ is congruent to $0, 1$, or -1 modulo one of ± 2 or ± 3. Hence θ or $\theta - 1$ or $\theta + 1$ is divisible by 2 or 3. But this is impossible since $N(\theta) = 5 = N(\theta - 1)$ and $N(\theta + 1) = 7$, while $N(2) = 4$ and $N(3) = 9$.

***A is a* PID.** As stated in the introduction, to show that A is a principal ideal domain (PID) it is enough to show that A is "almost" a Euclidean domain. More precisely, it may be seen that given elements $\alpha, \beta \in A$, $\beta \neq 0$, if β does not divide α and $N(\alpha) \geqslant N(\beta)$ then there exist γ, $\delta \in A$ such that

$$0 < N(\alpha\gamma - \beta\delta) < N(\beta). \tag{1}$$

This property implies that A is a PID by an argument similar to the one usually applied to show that $\mathbb{Z}$ is a PID. Let $I \neq 0$ be an ideal in A. Let $\beta \in I$ be an element such that $N(\beta)$ is minimal among the nonzero elements in I. Then $\beta A = I$. Indeed, since clearly $\beta A \subseteq I$, consider the possibility of having an element $\alpha \in I$ such that β does not divide α. Then $\alpha \neq 0$ and hence $N(\alpha) \geqslant N(\beta)$. Now using (i) it is possible to obtain another nonzero element $\alpha\gamma - \beta\delta$ in I which contradicts the minimality of $N(\beta)$.

To show (i) take $\alpha, \beta \in A$, $\beta \neq 0$. If β does not divide α and $N(\alpha) \geqslant N(\beta)$ write

$$\frac{\alpha}{\beta} = a + b\theta,$$

where a and b are rational numbers and at least one of them is not an integer. This is possible since the inverse of β as a complex number is in $\mathbb{Q}[\theta]$, which is a subfield of $\mathbb{C}$.

A case by case consideration leads to elements γ and $\delta \in A$ such that

$$0 < N\left(\frac{\alpha}{\beta}\gamma - \delta\right) < 1, \quad \text{whence} \quad N(\alpha\gamma - \delta\beta) < N(\beta).$$

There are seven cases.

Case 1: $b \in \mathbb{Z}$. Then $a \notin \mathbb{Z}$ and we may take $\gamma = 1$ and $\delta = \{a\} + b\theta$ (here $\{x\}$ denotes the integer nearest x, with $\{n + 1/2\} = n$). Now,

$$0 < N\left(\frac{\alpha}{\beta}\gamma - \delta\right) \leqslant \frac{1}{4} < 1.$$

Case 2(a): $a \in \mathbb{Z}$ and $5b \notin \mathbb{Z}$. Then $\dfrac{\alpha}{\beta}\bar{\theta} = a + 5b - a\theta$ and we may take $\gamma = \bar{\theta}$, $\delta = \{a + 5b\} - a\theta$.

Case 2(b): $a \in \mathbb{Z}$ and $5b$ $\mathbb{Z}$. Take $\gamma = 1$, $\delta = a + \{b\}\theta$.

Case 3(a): $a,b \notin \mathbb{Z}$ and $2a, 2b \in \mathbb{Z}$. Then, although we proved IV for a, $b \in \mathbb{Z}$, it is clearly valid also for a, b rational and hence $\theta\alpha/\beta = -5b + (a + b)\theta$ and $a + b \in \mathbb{Z}$. Therefore, we may take $\gamma = \theta$, $\delta = \{-5b\} + \{a + b\}\theta$.

Case 3(b): $a, b \notin \mathbb{Z}$ *and* $2a, 2b \notin \mathbb{Z}$. Then either $|b - \{b\}| \leqslant 1/3$ or $|2b - \{2b\}| \leqslant 1/3$. In the first situation take $\gamma = 1$ and $\delta = \{a\} + \{b\}\theta$ and estimate

$$0 < N\left(\frac{\alpha}{\beta}\gamma - \delta\right) \leqslant \frac{35}{36} < 1.$$

In the second situation take $\gamma = 2$ and $\delta = \{2a\} + \{2b\}\theta$ with the same estimate.

Case 3(*c*): $a, b \notin \mathbb{Z}$, $2a \in \mathbb{Z}$ and $2b \notin \mathbb{Z}$. When $5b \in \mathbb{Z}$ take $\gamma = 5$ and $\delta = \{5a\} + 5b\theta$ and when $5b \notin \mathbb{Z}$ take $\gamma = 2\bar{\theta}$ and $\delta = \{2a + 10b\} - 2a\theta$.

Case 3(*d*): $a, b \notin \mathbb{Z}$, $2b \in \mathbb{Z}$ and $2a \notin \mathbb{Z}$. Take $\gamma = 2$, $\delta = \{2a\} + 2b\theta$.

REFERENCES

1. J. Goldhaber and G. Ehrlich, Algebra, Collier Macmillan, London, 1970.
2. T. Hungerford, Algebra, Springer-Verlag, New York, 1974.
3. N. Jacobson, Lectures in Abstract Algebra, Van Nostrand Company Inc., Toronto, 1951.
4. S. Lang, Algebra, Addison-Wesley, Reading, 1965.
5. T. Motzkin, The Euclidean algorithm, *Bull. Amer. Math. Soc.* 55 (1949) 1142–1146.

PROBLEMS AND SOLUTIONS

For instructions about submitting ***proposed*** *problems for publication in this department see the inside front cover. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

Solutions *should be sent to the address given on the inside front cover. Two copies suffice.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method ~~yields a~~ more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

For instructions about submitting solutions of Problems, which should be mailed before March 30, 1989, see the inside front cover. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label if an acknowledgement is desired.

ELEMENTARY PROBLEMS

E 3289. *Proposed by C. Alsina, Universitat Politecnica de Catalunya, Barcelona, Spain.*

Let $I = [0,1]$. Prove that there are exactly two mappings T from $I \times I$ to I which have the following five properties for all x, y, z in I:

(1) $T(T(x, y), z) = T(x, T(y, z))$,
(2) $T(x, y) = T(y, x)$,
(3) $T(x, z) \leqslant T(y, z)$ if $x \leqslant y$,
(4) $T(x, 1) = x$,
(5) $T(zx, zy) = z^k T(x, y)$ for some fixed positive k.

E 3290. *Proposed by Douglas B. West, University of Illinois, Urbana, IL, and Douglas H. Wiedemann, Institute for Defense Analyses, Princeton, NJ.*

Given nonnegative integers l, m, n with $l, n \leqslant m$, evaluate the double sum

$$\sum_{i=0}^{l} \sum_{j=0}^{i} (-1)^j \binom{m-i}{m-l} \binom{n}{j} \binom{m-n}{i-j}.$$

E 3291. *Proposed by Jerrold R. Griggs, University of South Carolina, Columbia, SC.*

Suppose $y > x > 0$. Is it always true that $y^{x^y} > x^{y^x}$?

E 3292. *Proposed by Solomon W. Golomb, University of Southern California, Los Angeles, CA.*

(a) What is the maximum number of consecutive days with no occurrences of Friday the thirteenth?

(b) What is the minimum number of consecutive days with four occurrences of Friday the thirteenth?

E 3293. *Proposed by Joseph Keane (student), Carnegie-Mellon University, Pittsburgh, PA, and Gregg Patruno, The First Boston Corporation, New York, NY.*

Suppose that the distinct circles C_1 and C_2 intersect at P and Q. Suppose that the tangent to C_1 at P intersects C_2 again at A, the tangent to C_2 at P intersects C_1 again at B, and the line AB separates P and Q. Let C_3 be the circle externally tangent to C_1, externally tangent to C_2, tangent to line AB, and lying on the same side of AB as Q. Prove that the circles C_1 and C_2 intercept equal segments on one of the tangents to C_3 through P.

E 3294. *Proposed by Daniel Goffinet, Saint Étienne, France.*

Characterize the topological spaces in which the following assertion is true: For each open subset G of a closed subset F there exists a set A with G as its interior and F as its closure.

SOLUTIONS OF ELEMENTARY PROBLEMS

Zeros of a Combination of Orthogonal Polynomials

E 2979 [1982, 757]. *Proposed by Emeric Deutsch, Polytechnic Institute of New York.*

Let $\{p_k\}$ and $\{q_k\}$ $(k = 0, 1, 2, \dots)$ be two sequences of real, monic, orthogonal polynomials, p_k and q_k having degree k. Show that for any nonnegative integers m and n, the polynomial $p_m q_{n+1} + p_{m+1} q_n$ has only real simple zeros.

Composite Solution based on the solutions of Richard Askey, University of Wisconsin (Madison), David Secrest, University of Illinois at Urbana-Champaign, and the proposer. By Theorems 3.3.1 and 3.3.2 of Gabor Szegö's *Orthogonal Polynomials* (Amer. Math. Soc., Providence, 1975) the zeros of p_k are real and simple, the zeros of p_m and p_{m+1} interlace, the zeros of q_k are real and simple, and the zeros of q_n and q_{n+1} interlace. It follows (cf. op. cit., Theorem 3.3.5) that we have partial fraction decompositions:

$$\frac{p_m(x)}{p_{m+1}(x)} = \sum_{j=0}^{m} \frac{a_j}{x - u_j}, \qquad \frac{q_n(x)}{q_{n+1}(x)} = \sum_{k=0}^{n} \frac{b_k}{x - v_k},$$

where $u_0, u_1, \dots, u_m$ are the zeros of p_{m+1}, $v_0, v_1, \dots, v_n$ are the zeros of q_{n+1}, and the coefficients a_j and b_k are *positive*.

Now let S be the given polynomial $p_m q_{n+1} + p_{m+1} q_n$ and put

$$F = \frac{S}{p_{m+1} q_{n+1}} = \frac{p_m}{p_{m+1}} + \frac{q_n}{q_{n+1}}.$$

Let r be the number of common zeros of p_{m+1} and q_{n+1}. Then F has exactly $m+n+2-r$ poles, all of which are real. Since both p_m/p_{m+1} and q_n/q_{n+1} have positive residues at their poles, it follows by addition that the residue of F at each of its poles is positive. Accordingly F goes from $+\infty$ to $-\infty$ on the interval between any two consecutive poles. Hence F must have at least $m+n+1-r$ zeros between its poles; these zeros are also zeros of S, since $S = Fp_{m+1}q_{n+1}$. However, since the poles of F are simple, the r common zeros of p_{m+1} and q_{n+1} are also zeros of S and are distinct from the above mentioned zeros of F. Hence S must have at least $m+n+1$ distinct real zeros; since $\deg S = m+n+1$, in fact S has exactly $m+n+1$ real simple zeros.

Trisecting a Triangle by Cevians

E 3155 [1986, 482]. *Proposed by Gene Bennett, John Glenn, and Clark Kimberling, University of Evansville.*

Prove that for any triangle ABC, there exist points A', B', C' satisfying
(1) A' lies on side $\overline{BC}$, B' on side $\overline{AC}$, and C' on side $\overline{AB}$.
(2) $\overline{A'C} + \overline{CB'} = \overline{B'A} + AC' = C'B + BA'$; and
(3) $\overline{AA'}$, $\overline{BB'}$, and $\overline{CC'}$ concur in a point.

Composite Solution based on the solutions of Jeffrey M. Cohen and others. Let $\overline{AB}$ be the shortest side. For any choice of C' on $\overline{AB}$, determine points A' and B' on the perimeter of triangle ABC which are one-third of the way around the perimeter from C' in the opposite direction from A and B respectively. (One of these will be on the longest side; the other one will be either on the third side or on the longest side.)

The location of the point I of intersection of $\overline{AA'}$ and $\overline{BB'}$ is a continuous function of C'. It lies on the B side of $\overline{CC'}$ when $C' = A$ and on the A side of $\overline{CC'}$ when $C' = B$. The Intermediate Value Theorem implies that, as C' sweeps through all positions from A to B, there must be at least one position where $\overline{CC'}$ contains I. When this happens, A' will be on the side $\overline{BC}$ and B' will be on the side $\overline{CA}$.

Editorial comment. Most solvers used the 17th-century theorem of Ceva from projective geometry. Suppose a line is drawn from each vertex of a triangle to a point on the opposite side, thus cutting each side into two segments. Ceva's theorem asserts that the product of the lengths of the three segments clockwise from the vertices equals the product of the lengths of the three segments counterclockwise from the vertices if and only if the three lines drawn are concurrent. When the lines are concurrent and the two triple products are equal, the three points on the sides are called "Cevian points." Interestingly, the solvers using Ceva's theorem were almost precisely those *not* from North America; Ceva's theorem seems much better known among mathematical communities outside North America. In the present instance the use of Ceva's Theorem was not an indispensable part of the argument.

Solved by J. M. Cohen, J. Dou (Spain), J. Fukuta (Japan), L. Kuipers (Switzerland), N. Lord (England), O. P. Lossers (The Netherlands), E. Morgantini (Italy), V. Pambuccian (Romania), A. Pedersen (Denmark), W. Raffke (West Germany), V. Schindler (East Germany), L. Smith (student, Canada), J. H. Steelman, P. Tzermias (Greece), P. J. Zwier, and the proposers. Two incorrect solutions were received.

Inside a Difference Set

E 3176 [1986, 733]. *Proposed by Anatole Beck, The London School of Economics and Political Science, England.*

Let A be a measurable subset of R. Is it true that the interior of $A - A$ (set of differences) either contains 0 or contains nothing?

Solution by O. P. Lossers, Eindhoven University of Technology, The Netherlands. The answer is "no." This can be shown as follows. Define

$$A_0 := \left\{ \sum_{n=1}^{\infty} \varepsilon_n 2^{1-2n} : \varepsilon_n \in \{0,1\} \text{ for all } n \right\},$$

$$A_1 := \left\{ 1 + \sum_{n=1}^{\infty} \varepsilon_n 2^{-2n} : \varepsilon_n \in \{0,1\} \text{ for all } n \right\},$$

$$A := A_0 \cup A_1.$$

Then A is obviously measurable (with measure zero). Since $A_0 = \{2/3\} - A_0$, $A_1 = \{7/3\} - A_1$, $A_0 + A_1 = [1,2]$, we obtain

$$A_0 - A_0 = A_0 + A_0 - \left\{\frac{2}{3}\right\} = \left\{ \sum_{n=1}^{\infty} \varepsilon_n 2^{1-2n} : \varepsilon_n \in \{0,1,2\} \text{ for all } n \right\} - \left\{\frac{2}{3}\right\},$$

$$A_1 - A_1 = A_1 + A_1 - \left\{\frac{7}{3}\right\} = \left\{ \sum_{n=1}^{\infty} \varepsilon_n 2^{-2n} : \varepsilon_n \in \{0,1,2\} \text{ for all } n \right\} - \left\{\frac{1}{3}\right\},$$

$$A_1 - A_0 = A_1 + A_0 - \left\{\frac{2}{3}\right\} = \left[\frac{1}{3}, \frac{4}{3}\right],$$

$$A_0 - A_1 = \left[-\frac{4}{3}, -\frac{1}{3}\right].$$

Now $A - A$ is the union of these four sets and so it has a non-empty interior. However, if $-1/3 < x < 1/3$ and both $x + 1/3$ and $x + 2/3$ have three consecutive ones in their binary expansion (e.g., if $x = 2^{-k} + 2^{-k-3}$, $k \geqslant 2$) then $x \notin A_i - A_i$ $(i = 0, 1)$, so $x \notin A - A$ and hence $0 \notin (A - A)^{\circ}$.

Also solved by O. Matouš (Czechoslovakia) and the proposer.

An Inequality for Polynomial Roots

E 3184 [1987, 71]. *Proposed by Calin P. Popescu, Bucharest, Romania.*

Suppose the roots of the polynomial $P(x) = \sum_{i=0}^{n} a_i x^i$, $n \geqslant 2$, are real numbers lying in $(0, 1)$ and that $a_n > 0$. Show that $\sum_{i=k}^{n-2} \binom{i}{k} a_i > 0$, where $0 \leqslant k \leqslant n - 2$.

Solution I by David Callan, University of Bridgeport, Bridgeport, CT. Without loss of generality, we may assume $a_n = 1$. Let $x_1, \ldots, x_n$ be the roots of P. We evaluate the kth derivative $P^{(k)}(1)$ in two ways. First, using the fact that $\sum x_i = -a_{n-1}$, we

obtain

$$P^{(k)}(1) = k! \sum_{i=k}^{n} a_k \binom{i}{k} 1^{i-k} = k! \left\{ \binom{n}{k} - \binom{n-1}{k} \sum_{j=1}^{n} x_j + T_k \right\},$$

where

$$T_k = \sum_{i=k}^{n-2} \binom{i}{k} a_i$$

is the desired sum. Alternatively we perform k differentiations of the expression $P(x) = \prod(x - x_j)$ and then substitute $x = 1$. To each term in the resulting sum we apply the inequality $\prod(1 - y_j) > 1 - \sum y_j$, which holds for any r numbers $y_1, \dots, y_r$ between 0 and 1 by induction on r. The resulting computation, in which the sum runs over the sets $\{j_1, j_2, \dots, j_{n-k}\}$ of $n - k$ indices between 1 and n inclusive, gives

$$P^{(k)}(1) = k! \sum \prod_{i=1}^{n-k} (1 - x_{j_i}) > k! \left\{ \binom{n}{k} - \binom{n-1}{n-k-1} \sum_{j=1}^{n} x_j \right\}.$$

Equating the two expressions for $P^{(k)}(1)$ yields $T_k > 0$.

Solution II by O. P. Lossers, Eindhoven University of Technology, Eindhoven, The Netherlands. Weierstrass's Inequality, easily proved by induction on n, states that $\prod_{j=1}^{n}(1 - x_j) > 1 - \sum_{j=1}^{n} x_j$ for any set of n numbers $x_1, \dots, x_n$ between 0 and 1. We apply this to the case $k = 0$. Letting x_i be the roots of $P(x)$, we have $P(x) = a_n \prod_{j=1}^{n}(x - x_j)$ and $\sum_{j=1}^{n} x_j = -a_{n-1}/a_n$. This yields

$$\sum_{i=0}^{n-2} \binom{i}{0} a_i = P(1) - a_{n-1} - a_n = a_n \left[\prod_{j=1}^{n} (1 - x_j) + \sum_{j=1}^{n} x_j - 1 \right] > 0.$$

For $k > 0$, the kth derivative of $P(x)$ is the polynomial $Q(x) = k! \sum_{i=k}^{n} a_i \binom{i}{k} x^{i-k}$ of degree $n - k$. The leading coefficient of $Q(x)$ is positive and, by Rolle's Theorem, its roots lie between 0 and 1. Therefore we can apply the argument above to $Q(x)$ to obtain $k! \sum_{i=k}^{n-2} a_i \binom{i}{k} > 0$.

Also solved by A. Aziz (India), J. Ferrer (Spain), A. Riese, J. H. Steelman, A. Stenger, Univ. of S. Alabama Problem Group, and the proposer.

A Coloring Problem for Subsets of a Set

E 3188 [1987, 72]. *Proposed by Ioan Tomescu, University of Bucharest, Romania.*

Let X be a nonempty set having n elements and C be a color set with $p \geqslant 1$ elements. Find the greatest number p satisfying the following property: If we color in an arbitrary way each subset of X with colors from C such that each subset receives only one color, then there exist two distinct subsets A, B of X such that the sets A, B, $A \cup B$, $A \cap B$ have the same color.

Solution by Rhodes Peele, University of South Alabama, Mobile, AL. The answer is $p = n$. If $|C| \leqslant n$, then in any coloring of the subsets of $X = \{x_1, \dots, x_n\}$ the

pigeonhole principle implies that some pair of distinct subsets in the chain $\varnothing$, $\{x_1\}, \{x_1, x_2\}, \ldots, X$ have the same color. For such a pair A, B with $A \subset B$, we have $A \cap B = A$ and $A \cup B = B$, so they form the required subconfiguration.

Conversely, suppose $|C| > n$, and color the subsets of X in such a way that A gets color $|A|$. If A and B have the same color, then $A \cap B$ and $A \cup B$ have other colors.

Editorial comment. Note that we are coloring the elements of the power set of X, not the elements of X. O. P. Lossers pointed out that if the four sets A, B, $A \cup B$ and $A \cap B$ are required to be distinct, then the maximum value of p can be much lower than n. In this case $A \cap B$ and $A \cup B$ must have different cardinalities from A and B, because $|A| + |B| = |A \cup B| + |A \cap B|$. When $n = 6$, giving A color 1 if $|A| \in \{0, 1, 4, 6\}$ and color 2 if $|A| \in \{2, 3, 5\}$ yields no monochromatic configuration; since color depends only on cardinality, such a configuration requires a monochromatic solution i, j, k, l to $k + l = i + j$ with $i < k \leqslant l < j$. Thus for $n = 6$, the maximum p is 1. More generally, the maximum value of p is less than the number of colors required to color $\{0, \ldots, n\}$ without having a monochromatic solution i, j, k, l to $k + l = i + j$ with $i < k \leqslant l < j$.

Also solved by R. E. Bernstein, N. Falsinger, J. Ferrer (Spain), O. P. Lossers (The Netherlands), J. H. Steelman, P. Tracy, and the proposer.

An Alternating Ratio Sum

E 3190 [1987, 181]. *Proposed by Vasanth B. Solomon, Drake University, Des Moines, IA.*

Show that

$$\sum_{r=0}^{j} \frac{(-1)^r (N - 2r)\binom{j}{r}}{(N - r) \cdots (N - r - j)} = 0 \quad \text{for } j > 0 \text{ and } N > 2j.$$

Solution by O. P. Lossers, Eindhoven University of Technology, Eindhoven, The Netherlands. The assertion is valid for all (complex) $N \notin \{0, 1, \ldots, 2j\}$. We begin with the partial fraction expansion

$$\left(\frac{1}{z}\right)\left(\frac{1}{z-1}\right) \cdots \left(\frac{1}{z-j}\right) = \frac{1}{j!} \sum_{t=0}^{j} (-1)^{j-t} \binom{j}{t} \frac{1}{z-t}.$$

Applying this with $z = N - r$, we can rewrite the desired summation as

$$\frac{(-1)^j}{j!} \sum_{r=0}^{j} \sum_{t=0}^{j} (-1)^{r+t} \binom{j}{r} \binom{j}{t} \frac{N - 2r}{N - r - t}.$$

Replacing $(N - 2r)/(N - r - t)$ by $1 + (t - r)/(N - r - t)$ breaks this expression into two double summations. The first is zero because $\sum_{t=0}^{j} (-1)^t \binom{j}{t} = 0$, and the second is zero because its summand is skew-symmetric in r and t.

Also solved by J. N. Fitch, I. Gessel, A. A. Jagers (The Netherlands), L. E. Mattics, B. E. Rhoades, J. H. Steelman, P. Tracy, K. H. Vijitha-Kumara, and the proposer.

A Characterization of Rapid Convergence

E 3201 [1987, 371]. *Proposed by Grahame Bennett, Indiana University, Bloomington, IN.*

Let us say that a convergent series of positive terms, Σa_n, is rapidly convergent if there exists a sequence of positive numbers $\{b_n\}$ such that

$$\sum_{n=N}^{\infty} a_n \sum_{m=1}^{n} b_m = O(b_N).$$

Prove that the series Σa_n is rapidly convergent if and only if

$$\sum_{n=N}^{\infty} a_n = O\left(\frac{1}{N}\right).$$

(This shows that the termwise sum of two rapidly convergent series is rapidly convergent.)

Solution by University of South Alabama Problem Group, Mobile, AL. Let $A_m = \Sigma_{k=m}^{\infty} a_k$ and $B_n = \Sigma_{k=1}^{n} b_k$. If Σa_k is rapidly convergent, then by definition there exists a number c such that

$$cb_n > \sum_{k=n}^{\infty} a_k B_k = A_n B_n + \sum_{k=n+1}^{\infty} A_k b_k$$

for all n. If $N \geqslant n$, restricting the sums yields $cb_n > A_N B_n + \Sigma_{k=n+1}^{N} b_k A_N = A_N B_N$. Summing this yields $cB_N > NA_N B_N$, and hence $A_N < c/N$.

Conversely, if $A_N \leqslant c/N$ for all N, set $b_n = n^{-1/2}$. Then $\Sigma_{n=N}^{\infty} a_n B_n = A_N B_N + \Sigma_{n=N+1}^{\infty} A_n n^{-1/2} < 2cN^{-1/2} + c\Sigma_{n=N+1}^{\infty} n^{-3/2} = O(N^{-1/2}) = O(b_N)$.

Also solved by M. S. MacPhail (Canada), N. Martin, and the proposer.

An Exact Difference Cover of Integers

E 3202 [1987, 372]. *Proposed by Paul Erdös, Hungarian Academy of Science, Budapest.*

Prove that there is an increasing sequence of positive integers $a_1, a_2, \ldots$ with the following two properties:

(1) Every positive integer n can be uniquely expressed in the form $n = a_j - a_i$,

(2) a_k/k^3 is bounded.

Solution by the proposer. Choose $a_1 = 1$, $a_2 = 2$. Suppose $a_1, a_2, \ldots, a_{2k}$ have already been chosen so that the $k(2k-1)$ differences $a_j - a_i$ with $1 \leqslant i < j \leqslant 2k$ are all distinct. Suppose m is the first positive integer that does not occur among these $k(2k-1)$ differences. We choose a_{2k+1} and a_{2k+2} so that $a_{2k+2} = a_{2k+1} + m$ and so that the $4k$ differences

$$\begin{array}{ll} a_{2k+1} - a_j & 1 \leqslant j \leqslant 2k \\ a_{2k+2} - a_j = a_{2k+1} + m - a_j & 1 \leqslant j \leqslant 2k \end{array}$$

differ from m and from the $k(2k-1)$ differences already obtained. In particular, a_{2k+1} must avoid the values $a_1, a_2, \ldots, a_{2k}$ and $4k[k(2k-1)+1]$ other values.

Choosing the smallest value not forbidden, we have

$$a_{2k+1} \leqslant 2k + 4k[k(2k-1)+1] + 1 = 8k^3 - 4k^2 + 6k + 1 \leqslant (2k+1)^3,$$

$$a_{2k+2} = a_{2k+1} + m \leqslant a_{2k+1} + k(2k-1) + 1 < (2k+2)^3.$$

The sequence so constructed begins $1, 2, 5, 7, 15, 22, \dots$.

Editorial Comment. Hang-Fai Yeung pointed out that this problem appeared as problem P290 in the *Canadian Mathematical Bulletin*, 23 (1980) 380, solved differently by Erdös in 24 (1981) 505. In the same *Bulletin*, 24 (1981) 497–499, Andrew Pollington and Charles Vanden Eynden showed it is possible to choose a_k in the interval $[ck^3, c(k+1)^3]$ for a suitable constant c.

Also solved by I. Kozma (Israel), L. E. Mattics, and H.-F. Yeung (Australia). Two incorrect solutions were received.

Uniquely Fibonacci

E 3210 [1987, 457]. *Proposed by Paul A. Smith (student), University of Würzburg, West Germany.*

Determine all pairs (m, n) of integers such that

$$1 \leqslant m \leqslant n, \qquad m^2 \equiv -1 \pmod{n}, \qquad n^2 \equiv -1 \pmod{m}.$$

Solution by students in the 1987 Mathematical Olympiad Program, U.S. Military Academy, West Point, NY. Let (a, b) be any pair of integers such that

$$1 \leqslant a \leqslant b, \qquad a^2 \equiv -1 \pmod{b}, \qquad b^2 \equiv -1 \pmod{a}, \qquad b > 1,$$

and consider the transformation $T(a, b) = (c, a)$, where $c = (a^2+1)/b$. Observe that $c^2 = (a^2+1)^2/b^2 \equiv -(a^2+1)^2 \equiv -1 \pmod{a}$ and $a^2 = bc - 1 \equiv -1 \pmod{c}$. Also, $1 \leqslant c \leqslant a$, with $c = a$ only when $a = c = 1$. Thus, we can repeatedly apply T through pairs satisfying the constraints and obtain $T^k(a, b) = (1, 1)$ for some $k > 0$. It now follows that $(a, b) = T^{-k}(1, 1)$, where $T^{-1}(m, n) = (n, (n^2+1)/m)$.

Let F_n denote the nth Fibonacci number. Since $T^{-1}(1,1) = (1,2) = (F_1, F_3)$ and $F_n^2 = F_{n-2}F_{n+2} + (-1)^n$ for $n \geqslant 3$, it follows by induction that $T^{-k}(1,1) = (F_{2k-1}, F_{2k+1})$ for all $k \geqslant 1$. Thus

$$(1,1), (1,2), (2,5), \dots, (F_{2k-1}, F_{2k+1}), \dots$$

constitutes the list of all pairs that satisfy the given condition.

Editorial Comment. Charles Aschbachér and Hugh Edgar independently noted that this problem was solved by James C. Owings, Jr., in "Solution of the System $a^2 \equiv -1 \pmod{b}$, $b^2 \equiv -1 \pmod{a}$," *Fibonacci Quarterly*, 25 (1987) 245–249.

The proposer remarked that the condition $m^2 \equiv -1 \pmod{n}$ implies that the Dedekind sum $s(m, n)$ is zero. (Cf. Hans Rademacher and Emil Grosswald, *Dedekind Sums*, Carus Monograph No. 16, 1972, page 26.) Similarly the condition $n^2 \equiv -1 \pmod{m}$ implies that $s(n, m) = 0$. By the reciprocity formula for Dedekind sums (op. cit., Chapter 2)

$$m^2 + n^2 + 1 = 3mn,$$

or

$$(3n - 2m)^2 - 5n^2 = -4.$$

Thus $(3n - 2m + n\sqrt{5})/2$ is a unit of norm -1 in the real quadratic field generated by $\sqrt{5}$, so that

$$\frac{3n - 2m + n\sqrt{5}}{2} = \left(\frac{1+\sqrt{5}}{2}\right)^{2j+1}$$

for some non-negative integer j. Hence

$$3n - 2m = \left(\frac{1+\sqrt{5}}{2}\right)^{2j+1} + \left(\frac{1-\sqrt{5}}{2}\right)^{2j+1},$$

$$n\sqrt{5} = \left(\frac{1+\sqrt{5}}{2}\right)^{2j+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{2j+1},$$

so that

$$m = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{2j-1} - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^{2j-1} = F_{2j-1},$$

$$n = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{2j+1} - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^{2j+1} = F_{2j+1}.$$

Also solved by the proposer and 27 other readers.

ADVANCED PROBLEMS

6583. *Proposed by Moshe Laub, Jerusalem, Israel.*

(a) Let A be the set of positive integers z such that z^2 is expressible as a sum of two nonzero squares. Prove that A has asymptotic density one.

(b) Let B be the set of positive integers z such that the triangular number $T_z = z(z+1)/2$ is expressible as a sum of two nonzero triangular numbers. Prove that B has asymptotic density one.

6584. *Proposed by E. T. Parker, University of Illinois at Urbana-Champaign.*

Suppose a finite group G satisfies the following condition: If H is a cyclic subgroup of G of prime-power order or if H is an elementary abelian subgroup of G, then H lies in the center of the normalizer of H in G.

(a) Show that G need not be abelian.

(b) Show that G must be nilpotent.

6585. *Proposed by Alexandru Lupaş, Sibiu, Romania.*

Prove that if $-1 < x < 1$, then

$$\sum_{k=1}^{n} \frac{\sin(k \arccos x)}{k} = \frac{\sqrt{1-x}}{2} \int_{-1}^{x} \frac{1 - P_n(y)}{1-y} \frac{dy}{\sqrt{x-y}},$$

where

$$P_n(x) = \frac{1}{2^n n!}\left(\frac{d}{dx}\right)^n (x^2 - 1)^n$$

is the nth Legendre polynomial.

SOLUTIONS OF ADVANCED PROBLEMS

6536 [1987, 194]. *Proposed by Pei Yuan Wu, National Chiao Tung University, Hsinchu, Taiwan.*

Let T be a bounded linear operator on a complex Hilbert space H such that $\dim\ker(T) < \infty$. Define

$$\alpha(T) = \inf_S \{\|T - S\|: \dim\ker(S) > \dim\ker(T)\}$$

and

$$\gamma(T) = \inf_x \{\|Tx\|: x \perp \ker(T) \text{ and } \|x\| = 1\}.$$

It is proved on p. 365 of J. B. Conway, *A Course in Functional Analysis* (Springer Verlag, New York, 1985) that $\gamma(T) \leqslant \alpha(T)$. Show that

$$\gamma(T) = \alpha(T).$$

Solution by John Henry Steelman, Indiana University of Pennsylvania. Given the result of Conway, it suffices to show that $\alpha(T) \leqslant \gamma(T)$. We show this to be true in any inner product space.

Given any $x \perp \ker(T)$ with $\|x\| = 1$, let M be the subspace spanned by x and $\ker(T)$. Since $H = M \oplus M^{\perp}$, one can define a bounded linear operator S by setting $S = 0$ on M and $S = T$ on $M^{\perp}$. Now any $y \in H$ can be written as $y = m + \lambda x + m'$ where $m \in \ker(T)$ and $m' \in M^{\perp}$. Then $(T - S)y = T(\lambda x) = \lambda Tx$, and by the generalized Pythagorean theorem,

$$\|y\|^2 = \|m\|^2 + \|\lambda x\|^2 + \|m'\|^2 \geqslant \lambda^2.$$

Hence

$$\frac{\|(T-S)y\|}{\|y\|} = \frac{\|\lambda Tx\|}{\|y\|} = \frac{|\lambda|}{\|y\|}\|Tx\| \leqslant \|Tx\|.$$

Thus $\|T - S\| \leqslant \|Tx\|$. Since $\dim\ker(S) > \dim\ker(T)$, it follows that $\alpha(T) \leqslant \gamma(T)$.

Scott Hochwald considers the polar decomposition $T = UP$ of T, and lets $P = 0 \oplus P_1$ with respect to $\ker(P) \oplus (\ker(P))^{\perp}$. He shows that if $\sigma(S)$ denotes the spectrum of the operator S then

$$\begin{aligned}\gamma(T) &= \inf\{\lambda: \lambda \in \sigma(P_1)\}\\ &= \left(\inf\{\lambda > 0: \lambda \in \sigma(T^*T)\}\right)^{1/2}.\end{aligned}$$

Also solved by Mordechai Falkowitz, Jesús Ferrer (Spain), Ramesh Garimella, Scott Hochwald, Eero Posti (Finland), University of South Alabama Problem Group, and the proposer.

6537 [1987, 195]. *Proposed by F. S. Cater, Portland State University, OR.*

Let (u_n) be a sequence of positive numbers. For each positive integer n and each $x \in (0,1)$, let $(2^n x) = 2^n x - [2^n x]$, where $[2^n x]$ denotes the greatest integer in $2^n x$. Let

$$A = \left\{x\colon 0 < x < 1 \text{ and } \liminf_{n\to\infty} u_n^{-1}(2^n x) = 0\right\}.$$

Prove that the measure of A is 0 or 1 according as Σu_n converges or diverges.

Editorial Comment. Jim Conklin observed that the solution of this problem can be obtained from Example 6.7, p. 85, of Patrick Billingsley's *Probability and Measure*, Second edition, Wiley, N.Y., 1986. It is also a corollary of Theorem 2A in Walter Philipp's article, Some metrical theorems in number theory, *Pacific J. Math.* 20 (1967) 109–127. In terms of probability theory Billingsley's Example 6.7 is a Borel-Cantelli lemma for weakly dependent events, whereas Philipp proves strong laws of large numbers for the indicator functions of such events. Here is how one can deduce the result from the above sources.

Suppose $\Sigma u_n = \infty$. Then, by a well known construction there exists a sequence $\varphi_n \downarrow 0$ such that $\Sigma u_n \varphi_n = \infty$ still holds. Now taking $x_n = u_n \phi_n$ in Billingsley or Philipp, for almost all x we have

$$(2^n x) \leqslant u_n \varphi_n \quad \text{for infinitely many } n.$$

Hence

$$\liminf u_n^{-1}(2^n x) \leqslant \lim \varphi_n = 0.$$

Now suppose $\Sigma u_n < \infty$. Construct a sequence $\psi_n \uparrow \infty$ such that $\Sigma u_n \psi_n < \infty$. By the same results as above, for almost all x we have

$$(2^n x) \leqslant u_n \psi_n \quad \text{for finitely many } n\text{'s only.}$$

Hence (for almost all x)

$$u_n^{-1}(2^n x) > \psi_n \quad \text{for all } n \geqslant N = N(x),$$

so that for almost all x we have

$$\lim u_n^{-1}(2^n x) = \infty.$$

Thus the measure of A is 0 or 1 according to whether $\Sigma u_n < \infty$ or $\Sigma u_n = \infty$.

Solutions were also received from L. E. Mattics and the proposer.

REVIEWS

EDITED BY JOSEPH KONHAUSER
Department of Mathematics, Macalester College, St. Paul, MN 55105

Superstring Theory. By Michael B. Green, John H. Schwarz, and Edward Witten. Cambridge Monographs on Mathematical Physics. Cambridge University Press, New York, 1987. Vol. 1, Introduction. x + 469 pp. Vol. 2, Loop Amplitudes, Anomalies and Phenomenology. xii + 596 pp.

JOEL A. SHAPIRO
Department of Physics and Astronomy, Rutgers University, Piscataway, NJ 08855

The front-page article of the *New York Times Magazine** called it "A Theory of Everything," while to the *Atlantic*† it was "The Gospel of String." Controversy over its appropriateness within the Halls of Physics has arisen in magazines ranging from *Physics Today*‡ to the *Chronicle of Higher Education*.§ The Times had previously announced¶: "There is a nagging suspicion abroad that the science of physics may be on its death bed."

What is this "Theory of Everything" (TOE), and why does it arouse such passion? It is a set of highly mathematical approaches, collectively called "String Theory," proposed as the totally unified theory of all the forces of Physics. This would make it the culminating step in the series of unifications which has marked the dramatic progress in fundamental Physics in recent times. While unification has a long history, its modern era began with the melding of electromagnetism with the "weak" force responsible for the radioactive decay of nuclei. The combined theory, called Salam-Weinberg or electroweak theory, is the first really successful gauge theory based on a non-Abelian symmetry group, $SU(2) \times U(1)$. The next unification step attempts to unify this Salam-Weinberg theory with Quantum Chromodynamics (QCD), the gauge field theory based on the color group $SU(3)$, which describes the strong interactions of hadrons. The resulting Grand Unified Theories (GUTS) require that at energy scales far beyond those available today, the underlying interactions are more symmetric, having at every point of space-time an enlarged symmetry group such as $SU(5)$ or $SO(10)$ or E_6.

These theories are all gauge field theories, which means they are built on a fiber bundle with a compact Lie structure group and space-time as the base manifold. The connection on this bundle, viewed as an operator on the space of quantum mechanical states of the system, describes the quantum field associated with a point particle of spin 1. For electromagnetism this gauge particle is the photon, and corresponds to the one generator of $U(1)$. In the electroweak theory, the $SU(2) \times U(1)$ generators give the $W^{\pm}$ and Z bosons in addition to the photon. QCD is the

*Oct. 18, 1987.

†April, 1986, p. 24.

‡P. Ginsparg and S. Glashow, "Desperately Seeking Superstrings?," *Physics Today*, May, 1986.

§Kim McDonald, "Controversial 'Superstring' Theory Captures Physicist's Imagination but Could Prove Impossible to Verify," July 23, 1986.

¶Malcolm W. Brown, "Physics May Be Up Against the Wall This Time," *The New York Times*, July 8, 1986.

theory of quarks and gluons, of which the hadrons are built. The quarks, which are not gauge particles, were invented first, but the eight generators of the $SU(3)$ are the gluons, the glue of the strong interactions. This glue is so strong that neither the gluons nor the quarks can escape from their composite particles, which are the protons, neutrons, pions, and others, that populate the hadronic world. When GUT models unify these into a larger group, the standard model's $SU(3) \times SU(2) \times U(1)$ becomes a subgroup of a larger group. The extra generators needed have particles (so far unobserved) called X-mesons or baseballs.

Gravity, the one remaining known interaction, is also a gauge field theory, but of a different type. In Einstein's general relativity theory, the gauge transformations are "general coordinate transformations," or diffeomorphisms, of the space-time manifold. The group is itself related to space-time, rather than representing an "internal" symmetry, and this causes the associated particle, the graviton, to have a higher spin. It is a spin-two, or tensor, particle. The quantum field theory of such particles does not provide an acceptable quantum mechanical description of gravity, because it is *nonrenormalizable*. Renormalization is the technique used in order to make sense of the problem that all four-dimensional (space + time) field theories, when evaluated by perturbation expansion about the noninteracting theory, produce infinite answers. These are due to the interactions taking place at the same point (ultraviolet singularities). In a renormalizable theory, it is possible to relegate the infinities into the unobservable parameters which describe the "bare" theory, expressing the resultant answers in terms of a finite number of finite "renormalized" coupling constants. But for gravity this is not possible.

String theory is now seen as a theory which does successfully quantize gravity, along with the other interactions. In string theory, the basic objects are not point particles tracing out curves as they travel through space-time, as they are in field theory—instead they are strings, open or closed, which sweep out a two-dimensional surface, called the world sheet, as they travel through space-time. The coupling of three strings takes place over a smooth surface with no distinguished point. As a consequence, the ultraviolet behavior is nonsingular. There arises a perturbative expansion for scattering amplitudes in which a given term of the perturbative expansion is a functional integral over maps of a Riemann surface $\mathscr{R}$ into space-time $\mathscr{M}$. For theories which involve open strings, the surfaces may have boundaries. The full perturbation expansion involves summing over terms, that is, over the genus of the Riemann surface. For open strings we also need to sum over the topology of the boundaries. For each term, we must integrate over the moduli, and over the embedding maps X^μ: $\mathscr{R} \to \mathscr{M}$. In addition to the map into space-time, the "super" models involve fermionic functions on the Riemann surface. These models have an enhanced symmetry, called supersymmetry, which is a Lie algebra with a $\boldsymbol{Z}_2$ grading. The particles corresponding to odd grade are fermions, such as electrons and quarks, which have half-integral spin.

The string picture is a great change from ordinary field theory, much more tightly constrained by requirements of consistency. Unlike ordinary field theory, in which various forms of interactions can be postulated, string theory offers no choice of interaction, because the interaction is purely geometrical, a topological feature of the embedded surface. Among the inescapable requirements of string theory is the requirement that the manifold through which the string moves have a particular (and large) dimension, 10 for superstrings and 26 for bosonic strings. There is also

no way to get around having a spin-two massless particle, and one generally gets a number of spin-one massless particles.

If superstring theory does turn out to be the TOE, historians of science will have a hard job explaining why it came into being. It originated in the model of Veneziano (1968), which described something completely different, the phenomenology of the scattering of mesons, most importantly the existence of a tower of unstable particles forming a linear plot in spin vs. mass-squared. The slope of this line defines a hadronic mass scale of about 1 GeV. After a dramatic proliferation of phenomenological papers, the field quickly changed emphasis in an attempt to make a consistent complete theory of hadronic physics. The outstanding problem of string theory in its hadronic phase (alias dual-resonance model) was the wrong dimensionality and the theory's insistence on having massless spin-one and spin-two particles, something not present in the hadronic world. Its lack of ultraviolet singularities also implied softness in deep inelastic scattering, just when hardness was emerging as the new experimental feature on which all theorists should focus their understanding. Finally, in 1974, it became accepted in the high-energy community that QCD provided the theory of hadrons, and interest in strings dropped dramatically. But just as the hadronic era of strings was collapsing, Scherk and Schwarz proposed that the same model, with its scale changed by a factor of 10^{19}, could describe a quantum theory of gravity. This idea did not fall on immediately fertile ground—during the period from 1976–1981, almost no work was done on strings. It was a period of work on GUTS, and also on supergravity, which is based on gauging supersymmetry. In 1977 it was noticed that one such theory is the massless sector of supersymmetric string theory. In the early '80s, Green and Schwarz developed the manifestly space-time supersymmetric form of string theory. Finally in 1984 they published their paper on anomaly cancellation, and almost overnight strings emerged from their long slumber and started the rise to the superstar status they currently enjoy.

Anomalies are problems which arise in theories with symmetries during the renormalization process. To regularize the formally divergent integrals, one may need to impose a cutoff which violates the symmetry. If the effect of this cutoff does not disappear into a renormalization, one has an anomaly, with consequences ranging from loss of symmetry in some theories, to inconsistencies in the case of gauge anomalies. Thus there was a big impact when it was shown that the gauge anomaly problems which plagued ordinary GUTS theories were solved in an almost unique way in string theory, provided the internal gauge symmetry group was one of two possibilities, $SO(32)$ or $E_8 \times E_8$. Furthermore, possible gravitational anomalies also cancelled under the same conditions. The excitement was enhanced because this was the first time a fairly abstract consistency requirement had picked out, almost uniquely, a symmetry group. At the time, only open string theories had gauge groups, and these could only be $SO(N)$ or $SP(N)$, so there was no $E_8 \times E_8$ string. Soon thereafter, however, the heterotic closed string was invented. It requires an even, self-dual, 16-dimensional lattice, again picking out $SO(32)$ and $E_8 \times E_8$! It is even said that $E_8 \times E_8$ has properties good for phenomenology.

In this second active period, since the anomaly cancellation, string-theory research has been distinguished by the level of mathematical sophistication of much of the development. The bias against formal mathematics, intensely instilled into physics students in the '60s, began to erode during the late '70s, when everyone was

exploring gauge theories. The power of notions of algebraic topology, such as index theorems, was proven in Yang-Mills theory. A greater appreciation of differential geometry came about because the compactifications attempted in supergravity involved some rather unintuitive Riemann spaces, so that more sophisticated mathematics was needed to investigate them. Still more than Yang-Mills theories or supergravity, however, strings provide a rich field for the mathematically inclined. Even the simplest theory, the bosonic closed string, has a sum over all embeddings of Riemann surfaces into space-time. This requires a measure on moduli space, which must make the singular contributions from the boundaries of moduli space consistent with the physical requirement of unitarity.

The bosonic theory is a rich development tool, but physics requires fermions as well, and this leads to the superstring. Actually, the inclusion of fermions is also required to get rid of one fatal flaw which the boson theory has—a particle which travels faster than light, called the tachyon. When fermions are included, it is possible to abolish the tachyon. One method of incorporating fermions is to include a fermionic field representing a space-time spinor defined on the world-sheet. This leads to the manifestly space-time supersymmetric theory known as the Green-Schwarz string. An alternative approach to an equivalent theory is to consider supersymmetry on the world-sheet. In fact, this Ramond-Neveu-Schwarz formulation of 1971 was the first place where the supersymmetry idea became popular. Here we have a supermanifold, with one or two Grassmann coordinates in addition to the two ordinary ones. Quantization involves summing over all functions from super Riemann surfaces into ordinary space-time. So now we need new mathematics, the study of supermoduli space.

There is also a great deal of old but sophisticated mathematics now necessary in doing string theory. I will give one example. Physics requires that something be done with the six dimensions that are not manifest to our eyes and watches. Generalizing an old idea of Kaluza and Klein, we now believe that the ten-dimensional manifold, while nearly flat and very large in four dimensions, is tightly curled up in the other six. Thus the ten-dimensional space-time is approximately $M^4 \times K$, where M^4 is four-dimensional Minkowski space and K is a compact six-dimensional manifold. A quantum theory of gravity has a natural scale of 10^{19} GeV in energy or 10^{-33} cm in distances. If we assume that the dimensions of K are on that scale, all current physics is very low energy indeed. Thus all particles we know of are essentially massless, corresponding to zero modes of differential operators on K. In general, these zero modes will only be there if there is a reason for them, which brings in the concepts of index theorems and cohomology classes. Massless particles exist because a nonzero index or a nonzero Betti number requires them.

Another physical value which seems to be zero compared to the natural scale is the cosmological constant, which may be thought of as the energy density of the vacuum. In any ordinary field theory without supersymmetry, it is naively predicted to be infinite, because the vacuum energy density is the most divergent of all quantities in quantum field theories. Supersymmetry arranges a cancellation between the contributions of bosons and fermions. If the cosmological constant is neither zero nor infinite, it would be naively expected to be ~ 1 in natural units, where Planck's constant, the speed of light, and Newton's gravitational constant are all set to one. A density of 1 in these units corresponds to 5×10^{93} g/cm^3, whereas the experimental upper bound on the cosmological constant is 10^{-29} g/cm^3,

undoubtably the worst finite error of any prediction in science. Even if we begin with a theory that is supersymmetric before compactification, a compactification which breaks all supersymmetry at the gravitational scale could not be consistent with this density. This means that some supersymmetry must survive compactification; the state which corresponds to the vacuum is invariant under some supersymmetry transformations. The supersymmetry parameter then needs to satisfy the equations of a massless spinor field, which on K means there must exist a covariantly constant spinor field η. Examining the spin connection, which is a gauge field of $SO(6) \sim SU(4)$ on the six-dimensional manifold K, we can ask about the holonomy. The existence of the covariantly constant spinor η implies the group fixes the spinor η, so the holonomy must lie in the remaining $SU(3)$. The covariant spinor defines a Kähler form $\bar{\eta}\Gamma_{ij}\eta$, where Γ's are defined as antisymmetrized products of the Dirac matrices on K. Thus K becomes a three-dimensional complex manifold, and the restriction to $SU(3)$ holonomy means that the first Chern class vanishes, and K is a Calabi-Yau manifold. Classification of three-dimensional Calabi-Yau manifolds is a growth industry these days.

Writing a textbook which can give a graduate student the ability to access the current string literature is a monumental task. *Superstring Theory* represents a major effort by three of the most significant contributors to the field. Two of the authors are primarily responsible for the developments which caused the rebirth, and the other has been the most prominent exponent of strings since then. They have designed a text for advanced graduate students and others familiar with basic quantum field theory, and they have done an excellent job in covering all the most important areas without assuming more than is reasonable of the student's preparation. It is clear that a great deal of thought has been given to the approach, trying to use moderate tools and develop new ones as needed. Adequate discussion of results from the first incarnation are included, while maintaining a modern approach in the overall development. In an attempt to keep the total to 1,000 pages, many of the demonstrations are overly sketchy, so that working through the details is an educational but demanding process. A student attempting to do so will surely find places where the help of an expert in the field is required. The use of conformal field theory techniques is also not as well developed as I would have liked, and exercises would have been very useful.

There are also many hot topics which are not covered. The last two years has been a period during which much of the research is attempting a broader and deeper fundamental understanding of the whole theory. Several new abstract approaches to reformulation are being pursued. For example, the role of general coordinate invariance in space-time is not transparent in the conventional approach, and might be better understood in string field theory. Such current topics, which may or may not eventually prove to help in understanding strings, cannot reasonably be covered in a text at this time. Thus it is appropriate that these volumes cover the essential fundamentals of the field widely and well. This is clearly *the book* on string theory.

Calculus with Analytic Geometry. By George F. Simmons. McGraw-Hill Book Company. New York, 1985. xiii + 950 pp. $46.95.

UNDERWOOD DUDLEY
Department of Mathematics and Computer Science, DePauw University, Greencastle, Indiana 46135

This is about calculus books, and there are seven important conclusions.

Let us go back to the beginning and look at the first calculus book, *Analyse des Infiniment Petits, pour L'intelligence des Lignes Courbes*, published in Paris in 1696 and written by Guillaume François Antoine L'Hospital, Marquis de Sainte-Mesme, Comte d'Entrement, Seigneur d'Orques, etc. To almost everyone, L'Hospital is just a name attached to L'Hospital's Rule and almost no one knows anything about him. His memory deserves better. He displayed mathematical talent early, solving a problem about cycloids at fifteen, and he was a lifelong lover and supporter of mathematics who, unfortunately, died young, at the age of 43. Also, L'Hospital was no mean mathematician. He published several papers in the journals of the day, solving various nontrivial problems. I know that I could not have found, as he did, the shape of a curve such that a body sliding down it exerts a normal force on it always equal to the weight of the body. Further, as Abraham Robinson has written,

> According to the testimony of his contemporaries, L'Hospital possessed a very attractive personality, being, among other things, modest and generous, two qualities which were not widespread among the mathematicians of his time.

His book was a huge success. There was a second edition in 1715, and there were commentaries written on it. I have the 1781 edition, with additions made by another author. Not many textbooks last almost 100 years. Birkhoff and MacLane is not yet 50 years old. L'Hospital's book is about differentials and their applications to curves and the style is exclusively geometrical. There are not many equations, but there are an awful lot of letters and pictures, just as I remember in my tenth grade geometry text. Mathematics was geometry then, and mathematicians were geometers.

There are many things not in the book. There are no sines or cosines, no exponentials or logarithms, only algebraic functions and algebraic curves. There are also no derivatives, only differentials. Here is L'Hospital's proof of L'Hospital's Rule, using differentials. In Figure 1, points on the graph of $f(x)/g(x)$ are found by dividing lengths of abscissas, except at a. But if you dx past a, the point on the graph will be df/dg. Since dx is infinitesimal, that ratio gives you the point at a. Is that not nice?

It took some time for calculus to become generally taught in colleges. Eventually it made it, and calculus textbooks began to appear in the nineteenth century. I have a copy of one, *Elements of the Differential and Integral Calculus*, by Elias Loomis, Ll. D., Professor of Natural Philosophy at Yale College. His calculus was first published in 1851, and my copy of it is the 1878 edition. It sold in excess of 25,000 copies, so it must reflect accurately the style and content of calculus teaching of the time. Just as with L'Hospital, the differential was the important idea. Loomis derived the formula for the differential of x^n with no use of the binomial expansion, ($d(xy)/xy = dx/x + dy/y$, $d(x^n)/x^n = dx/x + \cdots + dx/x$, add and simplify) and his proof of L'Hospital's Rule was short, simple, and clear, and also one which

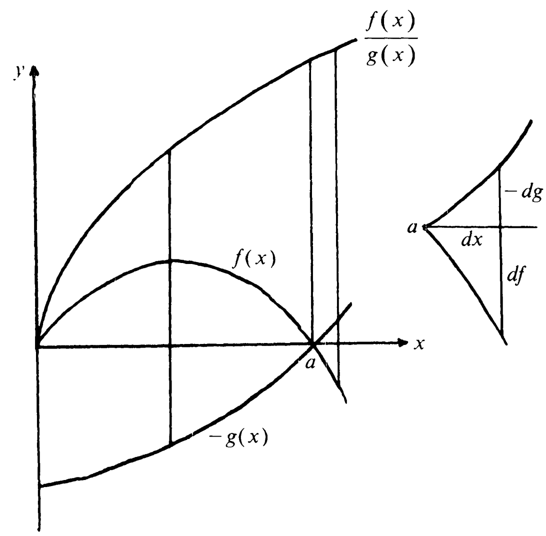


FIG. 1.

does not appear in modern texts because it fails for certain pathological examples. Also, Loomis put *all* of his formulas in words, italicized words. After deriving the formula for the differential of a power of a function, he wrote

> The differential of a function affected with any exponent whatever is the continued product of the exponent, the function itself with its exponent diminished by unity, and the differential of the function.

It was a good idea. It is also probably a good idea to do as L'Hospital and Loomis did and talk about differentials instead of derivatives whenever possible. Little bits of things are easier to understand than rates of change. It is a still better idea to strive for clarity and let students see what is really going on, which is what Loomis did, rather than putting rigor first. But nowadays authors cannot do that. They must protect themselves against some colleague snootily writing to the publisher, "Evidently Professor Blank is unaware that his so-called proof of L'Hospital's Rule is faulty, as the following well-known example shows. I could not possibly adopt a text with such a serious error." It is a shame, and probably inevitable that calculus books are written for calculus teachers, but I have nevertheless concluded

CONCLUSION #1: CALCULUS BOOKS SHOULD BE WRITTEN FOR STUDENTS.

It would be worth a try. *Calculus Made Easy* by Silvanus P. Thompson was quite successful in its time, which ran for quite a while. The second edition appeared in 1914, and my copy was printed in 1935. It is still in print. The book has a motto:

What one fool can do, another can.

and a prologue:

> Considering how many fools can calculate, it is surprising that it should be thought either a difficult or a tedious task for any other fool to learn how to master the same tricks. ...

> Being myself a remarkably stupid fellow, I have had to unteach myself the difficulties, and now beg to present to my fellow fools the parts that are not hard. Master these thoroughly, and the rest will follow. What one fool can do, another can.

Chapter 1, whose title is "To Deliver You From The Preliminary Terrors" forthrightly says that dx means "a little bit of x." Thompson did not include L'Hospital's Rule.

Both Loomis and Thompson are like L'Hospital when it comes to giving applications and examples: they are all almost entirely geometrical. Loomis's applications of maxima and minima are all about inscribing and circumscribing things, and so are all of Thompson's except one. In fact, all three books are full of geometry. Thompson concluded with arc length and curvature, Loomis had involutes and evolutes, cusps and multiple points, and lots of curve sketching. Did you know that the asymptote to $y^3 = x^3 + x^2$ is $y = x + 1/3$? I didn't until I read Loomis. It is nice to know. There must have been some reason why calculus books for more than 200 years taught so much geometry. Mathematics may no longer be synonymous with geometry, but we have discarded, wrongly I think, the wisdom of the ages, and I have concluded

CONCLUSION #2: CALCULUS BOOKS NEED MORE GEOMETRY.

Before writing this essay, I examined 85 separate and distinct calculus books. I looked at all of their prefaces, all of their applications of maxima and minima, and all of their treatments of L'Hospital's Rule. By the way, I found five different spellings of L'Hospital. There were the two you would expect, and Lhospital, as L'Hospital sometimes spelled his name. In addition, one author, not wanting to take chances, had it L'Hôspital, and one thought it was Le Hospital. Why are there so many calculus books, and why do they keep appearing? One could be cynical and say that the authors are all motivated by greed. But I do not think so. I think that authors write new calculus books because they have observed that students do not learn much from the old calculus books. Therefore, prospective authors think, "if I write a text and do things properly, students will be able to learn." They are wrong, all of them. The reason for that is

CONCLUSION #3: CALCULUS IS HARD.

Too hard, I think, to teach to college freshmen in the United States in the 1980s, but that is another topic.

If you plot the books' numbers of pages against their year of publication, you have a chart in which an ominous increasing trend is clear. The 1000-page barrier, first pierced in 1960, has been broken more and more often as time goes on. New highs on the calculus-page index are made almost yearly. Where will it all end? We can get an indication. The magic of modern statistics packages produces the least-squares line: Pages = 2.94 (Year) − 5180, showing that in the middle of the next millennium, the average calculus book will have 2,270 pages and the longest one, just published, will have 3,783 pages exclusive of index.

Why do we need 1000 pages to do what L'Hospital did in 234, Loomis in 309, Thompson in 301, and the text I learned calculus from, used exclusively for four

whole semesters, 14 semester-hours in all, in 416? There are several reasons. One, of course, is the large number of reviewers of prospective texts. No more can an editor make up his mind about the merits of a text, it has to go out to fifteen different people for opinions. And if one of them writes that the author has left out the $\tan(x/2)$ substitution in the section on techniques of integration, how can he or she do that, we won't be able to integrate $3/(4 + 5\sin 6x)$, how can anyone claim to know calculus who can't do that; isn't the easiest response to include the $\tan(x/2)$ substitution? Of course it is, in it goes, and in goes everything else that is in every other 1000-page text. It is impossible to escape

CONCLUSION #4: CALCULUS BOOKS ARE TOO LONG.

Another reason for the length is the current mania for Applications. If you go to *Books in Print* and look in the subject index under "Calculus" what you see is

The Usefulness of Calculus for the Behavioral, Life, and Managerial Sciences
Essentials of Calculus for Business, Economics, Life Sciences, and Social Sciences

and many, many similar titles. Now authors have to explain, with examples, what marginal revenue is, and consumer surplus, and what tracheae are whereas in the old days, all their readers knew what a cone was. A third reason is the supposed need to be rigorous. Now we see statements of L'Hospital's Rule that take up half a page and proofs of it that go on for three pages. My 416-page calculus book never even mentioned L'Hospital's Rule, and I never felt the lack. Its author never proved that the derivative of x^e was ex^{e-1}, but I was willing to believe it. Trying to include everything and trying to prove everything makes for long books. Everything gets longer. Prefaces used to be short, a page or less. Now they are five and six pages, hard sells for the incredible virtues of the text that follows, full of thanks to reviewers, to five or six editors, to wives, to students, even to cats.

Let me return to "applications." There aren't many, you know. In the 85 calculus books I examined, almost all of them had the Norman window problem—the rectangle surmounted by a semicircle, fixed perimeter, maximize the area. The semicircle always "surmounts." This is the sole surviving use of "surmounted" in the English language, except for the silo, a cylinder surmounted by a hemisphere. Only one author had the courage to say that the window was a semicircle on top of a rectangle. All the books had the box made by cutting the corners out of a flat sheet, all have the ladder sliding down the wall, all had the conical tank with changing height of water, all had the tin can with fixed surface area and maximum volume, all had the V-shaped trough, all had the field to fence, with or without a river flowing (in a dead straight line) along one side, all had the wire—usually wire, but sometimes string—cut into two pieces to be formed into a circle and a square, though some daring authors made circles and equilateral triangles. There are only finitely many calculus problems, and their number is *very* finite.

"Applications" are so phony. Ladders do not slide down walls with the base moving away from the wall at a constant rate. Authors know the applications are phony. One book has the base of the ladder sliding away from the wall at a rate of 2 feet per *minute*. At that rate, you could finish up your painting with time to spare and easily step off the ladder when it was a foot from the ground. Another author

has the old run-and-swim problem—you know, minimize the time to get somewhere on the other side of the river—with the person able to run 25 feet per second and swim 20 feet per second. That's not bad for running (it's a 3:31.2 mile), but it is super swimming, 100 yards in 15 seconds, a new world's record by far. There are no conical reservoirs outside of calculus books. Real reservoirs are cylindrical, or perhaps rectangular. The reason for this is found in the texts: in the problems, the conical reservoirs usually have a leak at the bottom. Tin cans are not made to minimize surface area. I could give any number of examples of absurd applications in which businessmen "observe" the price of their product decreasing at the rate of \$1 per month, or where the S. D. S. (remember them?) "find" that staging x demonstrations costs $\$250x^3$. Why will authors not be honest and say that these artificial problems provide valuable practice in translating from English into mathematics and that is all they are for? Surely they cannot disagree with

CONCLUSION #5: FIRST-SEMESTER CALCULUS HAS *NO* APPLICATIONS.

Before getting to my next conclusion, here is my favorite "application."

> A cow has 90 feet of fence to make a rectangular pasture. She has the use of a cliff for one side. She decides to leave a 10 foot gap in the fence in case the grass should get greener on the other side. Find ... .

Hardly any authors dare to do that. Calculus books are Serious. The text from which that problem came was titled *Calculus Without Analytic Geometry* and it is no surprise that it did not catch on.

The existence of all those calculus books with "Applications" in their titles implies a market for them. There must be students out there who are being forced to undergo a semester of calculus before they can complete their major in botany and take over the family flower shop. I cannot believe that any more than a tiny fraction of them will ever see a derivative again, or need one. Calculus is a splendid screen for screening out dummies, but it also screens out perfectly intelligent people who find it difficult to deal with quantities. I don't know about you, but I long ago concluded

CONCLUSION #6: NOT EVERYONE NEEDS TO LEARN CALCULUS.

The book by Simmons is a fine one. It was written with care and intelligence. It has good problems, and the historical material is almost a course in the history of mathematics. It is nicely printed, well bound, and expensive. Future historians of mathematics will look back on it and say, "Yes, that is an excellent example of a late twentieth-century calculus book." This leads to my last conclusion

CONCLUSION #7: THAT'S *ENOUGH* ABOUT CALCULUS BOOKS.

Geometry I. By Marcel Berger. Translated from the French by M. Cole and S. Levy. Springer-Verlag, New York, Berlin, Heidelberg, 1987, xiii + 427 pp.

Geometry II. By Marcel Berger. Translated from the French by M. Cole and S. Levy. Springer-Verlag, New York, Berlin, Heidelberg, 1987, x + 405 pp.

F. A. SHERK
Department of Mathematics, University of Toronto, Toronto, Ontario, Canada M5S 1A1

It would be an interesting study to follow the fortunes of geometry as a discipline within the broader framework of all mathematics. There was a time when geometry was dominant: every person with pretensions to being educated had to know geometry, and Euclid was king. This was true also in the mathematical world, where great attention was paid to Euclid. Attempts over many centuries to prove the Parallel Postulate led in the nineteenth century to the development of non-Euclidean geometry, while at about the same time a closer, more sophisticated critique of the foundations of Euclidean geometry laid the basis for the development of projective and affine geometry. On the latter, all the geometries of relevance to physics—Riemannian, Minkowski, and the like—are based. Spherical and inversive geometries were also developed, the result of all this activity being that geometry, if not supreme, was at least still a coruler in mathematics.

But as the present century proceeded, interest in geometry fell off; it began no longer to be classified in the "mainstream" of mathematics. Today, in North America at least, an undergraduate student majoring in mathematics will be lucky to have even one geometry course included in his or her program. High school has provided the student with the rudiments of Euclidean geometry—synthetic and analytic—and there may be some offerings of differential geometry in college. But it is rare to find other college undergraduate geometry courses, and still more rare to find substantial enrollments in those that do exist. The geometer on the staff, if indeed there is one, is swamped by analysts, topologists, and algebraists.

Yet we optimistic geometers sense a revival of interest in the subject. Certainly an impressive amount of geometry has been produced in the twentieth century, and continues to appear. Work on geometric transformations, for example, begun a century ago, has been developed beyond the wildest expectations of Klein's "Erlangen Programme," and well beyond the bounds that it intrinsically set. Not only has this given impetus to the abstract study of groups, by providing examples of many of the most important, but it has at times shown the way to better and shorter proofs of purely group-theoretic theorems. For example, the proof that the group $[p, q]^+ \cong \langle R, S \mid R^p = S^q = (RS)^2 = 1 \rangle$ is finite if and only if $(p-2)(q-2) < 4$ (p, q being positive integers) is long and tedious if dealt with abstractly. But Coxeter gave a fine proof in a few lines, by identifying $[p, q]^+$ with the rotational symmetry group of the regular tessellation $\{p, q\}$. Again, Coxeter's important work on groups generated by reflections has not only given clarity and precision to polytope theory, but it has provided a classification of an important family of discrete groups and given an excellent notation for them (the "Coxeter-Dynkin diagram").

The study of non-Desarguesian geometry, almost entirely a twentieth-century subject, is very interesting for its own sake. But it has also made many contributions

to the better understanding of nonassociative algebra, group theory, and combinatorial theory.

While the geometry studied today is often generalized to the point that it is based on any field (or even on a more general algebraic system, as in the case of non-Desarguesian geometry) there is still much to do in geometry over the real field. Convexity problems, extremal problems (such as close-packing of spheres), and tilings come readily to mind as examples. And, of course, there are many aspects of classical differential geometry that still need exploration. In all of these problems, modern analysis and topology become indispensable tools, both aiding the development of geometry and in turn being enhanced and clarified by their application to geometry.

One reason for optimism on the future of geometry is the appearance of good expository books on the subject. The two-volume work under review here is one of these. To quote from the Introduction, the author has three main objects in view:

—to emphasize the visual, or "artistic," aspect of geometry, by using figures in abundance;
—to accompany each new notion with as interesting a result as possible, preferably one with a simple statement but a non-obvious proof;
—finally, to show that this simple-looking mathematics does not belong in a museum, that it is an everyday tool in advanced mathematical research, and that occasionally one encounters unsolved problems at even the most elementary level.

In pursuing these objectives, the author naturally confines his attention to geometry over a field, with heavy emphasis on the field of real numbers. In that sense, the book deals with classical geometry, although many of the results dealt with are quite recent. It is a book which can be read by anybody with a college mathematical background, and could be used as a text for a number of different geometry courses. To the extent that that is possible, it need not be read in order from first chapter to last; rather, one can choose portions from various places in the book, with a reasonable chance of understanding the material. Some of the topics dealt with in the twenty chapters are geometric transformations and tilings, projective, affine, and Euclidean spaces, convex sets, polytopes, quadratic forms and orthogonality, spheres, elliptic and hyperbolic geometry.

There are many features of the content and style that are admirable, and one or two that are not. The selection of material—it being impossible to include everything—is consistent with the objectives quoted above, thus adding real excitement to the content without unduly skewing its systematic theoretical development. Definitions are immediately followed by several examples, and there are many exercises given to close off each topic. Each volume has an extensive bibliography, the contents of which are constantly referred to in the text. This feature, of course, has dual advantages of putting the topic under discussion into a historical context, and providing the serious student with easy access to further research.

But the most admirable feature of all, in this reviewer's opinion, is the astounding number of illustrations. These range from simple diagrams, put in exactly the right place relative to the text, to reproductions of art, pictures of mechanical gadgets, charts and maps. The ratio of illustrations to total pages of text is in fact slightly

greater than one. If this signals a trend to the return of pictures to help in the understanding of abstruse mathematical points, we should welcome it with becoming enthusiasm.

On the negative side, the literary style of the book leaves much to be desired. Particularly noticeable is the rather annoying occasional insertion of colloquial, folksy expressions ("If you don't know any exterior algebra..."; "The moral of the story is..."; "Here's a simpler way..."). This may, of course, be an indiscretion of the translators rather than the author. In any case, some mathematicians regard it as desirable to communicate in this light-hearted way; given that mathematics is serious business, requiring much concentration, it is not very appropriate in this reviewer's opinion. It is bad enough to have this style common in the classroom today; perhaps we should try to use something more appropriate in our serious textbooks.

The good features of the book, however, far outweigh the questionable. The work deserves a long run of popularity. And it will certainly help, rather than impede, the fortunes of geometry in its long trail back to the recognition it so richly deserves.

Pi and the AGM. By Jonathan M. Borwein and Peter B. Borwein. John Wiley & Sons, Inc., New York, Toronto, 1986, xv + 414 pp.

RICHARD ASKEY
Department of Mathematics, University of Wisconsin, Madison, WI 53706

The Mathematical Association of America, The American Mathematical Society and the Society for Industrial and Applied Mathematics have joined to try to get better media coverage of mathematics. One result of this is stories that say that computers have calculated 10 million, or now 130 million digits of π. When asked to comment on this by friends, most mathematicians wince and say they have no idea why anyone would want to do these calculations. Privately, they start to have serious doubts about more publicity for mathematics. If the computation of π to millions of places is to be the image of a mathematician, then it might be better to keep mathematics out of the newspaper. However, ten million is so much larger then the 707 digits that was claimed to be known when I first heard of this problem that something interesting must lie behind these calculations. Part of the improvement comes from larger and faster machines, but much more comes from increased mathematical knowledge. Part of this is specific to certain numbers, and part is general and of wide applicability. All of it is very interesting.

Calculations of π (really approximations to π, but I will be short and sloppy) really started with Archimedes. He used inscribed and circumscribed regular $6 \cdot 2^n$-gons and was able to show that $3\frac{10}{17} < \pi < 3\frac{1}{7}$ from $n = 4$. By the year 1600 or so this method had led to 34 digits. The next major advance was the arctangent series. While this was first discovered in India it was Gregory's work, which was used by Sharp under Halley's direction to obtain 71 digits. Machin used

$$\frac{\pi}{4} = 4\tan^{-1}\left(\frac{1}{5}\right) - \tan^{-1}\left(\frac{1}{239}\right)$$

to get 100 digits in 1706. W. Shanks eventually extended this type of calculation to 707 digits, 526 of which were correct. This calculation took years and his major

publication was in 1853. Electronic computers were used to reach one million digits in 1973, and the series used was similar to Machin's formula. In 1976 the mathematics used jumped to late 18th- and early 19th-century mathematics with the independent discovery by Salamin and Brent that the arithmetic-geometric mean algorithm of Lagrange and Gauss could be combined with an identity of Legendre to lead to a quadratically convergent algorithm.

Archimedes' algorithm can be stated as

$$a_{n+1} = \frac{2a_n b_n}{a_n + b_n}, \qquad b_{n+1} = (a_{n+1}b_n)^{1/2} \tag{A}$$

$a_0 = 2\sqrt{3}$, $b_0 = 3$.

The arithmetic-geometric mean algorithm is

$$a_{n+1} = \frac{a_n + b_n}{2}, \qquad b_{n+1} = (a_n b_n)^{1/2}. \tag{G}$$

If you have never seen algorithm (G) take out your pocket calculator and try both of these algorithms for $a_0 = 1$, $b_0 = 2$. The results for $n = 3$ should be a real surprise. They were for Gauss, and he was led to an integral representation for the limit in (G), and from this to elliptic functions. Lagrange had the integral first and discovered the arithmetic mean as a method of evaluating this integral. Since this solved the question he had, he did not push it further. This is an important point to remember, living as we do in a period when mathematicians are being encouraged to solve practical problems. Lagrange solved a practical problem, and quit after solving it. Gauss was playing around with numbers and recognized a number that arose as one he had calculated by another method when solving a practical problem. This problem was to evaluate $\int_0^{\pi/2}(a^2\cos^2\theta + b^2\sin^2\theta)^{-1/2}\,d\theta$. Gauss was struck by the surprising result that there was a way of representing the limit of (G) in terms of this integral, and by the incredible rapidity of the convergence of (G), and so examined his discovery in more detail. What he then discovered was much more important than the solution of how to evaluate the above integral in a rapid way. It was elliptic functions, as opposed to elliptic integrals, and he also discovered theta functions and the start of modular forms. (See the expository paper by Cox [**1**] for more on this.) I am not suggesting we not solve practical problems, but spend all our time experimenting with numbers; however, some time spent playing games is appropriate. Also, do not quit looking at a discovery after solving a problem. This discovery may have done much more than just solve the problem at hand.

In the fall of 1985 a contemporary computer wizard, Bill Gosper, decided to get in the competition for the largest calculation of π. He had two reasons. One was a desire to change the problem. The natural, invariant way to represent a real number is by a continued fraction rather than by a decimal or binary expansion. He hoped to convince others of this. The second was a request from his employers at Symbolics that he show that their relatively small LISP machine, a general computer that was not especially built for numerical work, could be used to compete with giant machines in the area for which large machines were designed—large scale numerical computations. Gosper was successful in getting 17 million digits, a record for a short while, and also the corresponding continued fraction. As frequently happens, he discovered a subtle design error in the computer. This is one reason massive calculations that can be checked are important.

The people using large machines were using nineteenth-century mathematics. Gosper introduced twentieth-century mathematics when he used the following result of Ramanujan:

$$\frac{1}{\pi} = 2\sqrt{2}\sum_{n=0}^{\infty}\frac{\left(\frac{1}{4}\right)_n\left(\frac{1}{2}\right)_n\left(\frac{3}{4}\right)_n}{(1)_n(1)_n n!}(1103 + 26390n)\left(\frac{1}{99}\right)^{4n+2} \tag{R}$$

with $(a)_n = a(a+1)\cdots(a+n-1)$, $n = 1,2,\ldots,(a)_0 = 1$. Gosper asked if I knew how to prove (R), and I had to admit I did not. Ramanujan had not given a proof. While the convergence in (R) is linear, it is so rapid that the crossover point when the quadratic methods become better is surprisingly high. Gosper was unsuccessful in getting others to compute continued fractions, so while 132 million digits of π are now available, the largest continued fraction calculation is still Gosper's.

In the nineteenth-century calculations of π the arithmetic was done without any sophisticated methods. It is now necessary to be very sophisticated and use something like the fast Fourier transform to do multiplication.

If this subject begins to sound more interesting than it did in the last newspaper article on 130 million digits of π, I have partly succeeded. To succeed completely I will have gotten you interested enough to read the delightful and important book by the Borweins with the strange title given above. Start with the last chapter, which gives a history of the calculations of π, then includes some exercises that give the irrationality of e and π and also their transcendence. Enough hints are given to make it possible for many people to carry out the details. Finally, irrationality measures for π, $\zeta(2) = \pi^2/6$, and $\zeta(3)$ are given using Beukers' version of Apéry's proof of the irrationality of $\zeta(3)$. Then start at the beginning and work your way through a portion of such topics as the arithmetic-geometric mean, elliptic integrals, theta functions, elliptic modular functions, modular equations, hypergeometric functions, computational complexity, the fast Fourier transform, other means and their iteration, lattice point sums, and other topics. As one might expect there are more than twice as many references to Ramanujan than to anyone else, since Ramanujan's work on modular equations was significantly deeper than that of others. It is fitting in the year of Ramanujan's centenary to realize that he was really a very modern mathematician. Surprisingly, the work of his that is mentioned most in this book comes from a paper he wrote in India before coming to England. The paper was rewritten after Ramanujan arrived in England so the appropriate references could be added, but Hardy wrote that all the work had been done while Ramanujan was in India. We are only now realizing what a remarkable paper this is.

If you are curious about how to prove (R), an outline is given in Chapter 5. Also, the next time you read a newspaper story about mathematics, read it the way political news needs to be read in many parts of the world, that is, read between the lines to get the story that is really there. It is probably very interesting, unlike the story that appeared in the paper. The problem of educating a few reporters so that they appreciate and understand some mathematics can probably be solved, but the problem of educating the editors to realize that not only does mathematics matter, but they should treat it seriously is much harder. However, an existence theorem exists at the *Los Angeles Times*, both for a reporter and an editor.

REFERENCE

1. D. A. Cox, The arithmetic-geometric mean of Gauss, *L'Enseignement Math.*, 30 (1984) 275–330.

TELEGRAPHIC REVIEWS

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T: Textbook	P: Professional Reading	1–4: Semesters
C: Computer Software	L: Undergraduate Library	**: Special Emphasis
S: Supplementary Reading	13: Grade Level	??: Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

General, S(10-16), P, L**.** *Archimedes' Revenge: The Joys and Perils of Mathematics.* Paul Hoffman. WW Norton, 1988, viii + 285 pp, $17.95. [ISBN: 0-393-02522-5] A vivid, engaging account of selected challenges of modern mathematics—primes, tiling, parallel computation, voting theory—as told through the actions and words of individuals who work on these problems: students, mathematicians, programmers, scientists, amateurs. Author Hoffman, editor of *Discover* magazine, writes in a lively, elementary style suitable for readers at all levels—from teenagers to grandparents. LAS

General, S*, P, L.** *Rubik's Cubic Compendium.* Ernö Rubik, *et al.* Transl: David Singmaster. Recreations in Math., V. 3. Oxford U Pr, 1987, xi + 225 pp, $26.95. [ISBN: 0-19-853202-4] Translation of a volume first published in 1981 in Hungarian with an "afterword" by translator-cubist Singmaster featuring the numerous Rubik "offspring" that emerged during the intervening years. Essays in the volume cover the mathematics, the psychology, the art, and the "play" of the cube—the latter by inventor Rubik himself. LAS

General, S. *Dr. Z's Beat the Racetrack, Revised Edition.* William T. Ziemba, Donald B. Hausch. William Morrow, 1987, 524 pp, $22.50. [ISBN: 0-688-07221-6] Updated version of the authors' 1984 *First Edition* (TR, August-September 1985), which outlined a winning strategy based on an inefficiency in the place- and show-betting markets. Includes many more exámples where the Dr. Z system was used to advantage. RSK

Education, P, L*. *Mathematics Education in Secondary Schools and Two-Year Colleges: A Sourcebook.* Paul J. Campbell, Louise S. Grinstein. Garland, 1988, xvii + 439 pp. [ISBN: 0-8240-8522-1] An annotated bibliography of books and articles suitable as references for teachers in high school and two-year colleges, organized into 20 sections on topics such as testing and evaluation, statistics and probability, slow learners, and teacher education. Brief essays by specialists open each section and provide context for the bibliography; author and subject indices help locate items. LAS

Education, P*, L*. *Guidelines for the Continuing Mathematical Education of Teachers.* Committee on the Mathematical Education of Teachers (COMET). MAA Notes No. 10. MAA, 1988, ix + 83 pp, (P). [ISBN: 0-88385-060-5] Detailed, practical recommendations for inservice and master's degree courses for practicing teachers of mathematics with sage boldfaced advice ("Teach a little well, rather than a lot badly"), and extensive references (over half the volume) for elementary, middle school, and secondary school teachers. A timely, important document that should help stimulate high-quality teaching at all levels. LAS

History, T(13-18: 1), S, L.** *A Concise History of Mathematics, Fourth Revised Edition.* Dirk J. Struik. Dover, 1987, xii + 228 pp, $7.95 (P). [ISBN: 0-486-60255-9] A new chapter on the mathematics of the first half of the twentieth century has been added to this edition (*Third Edition,* TR, November 1967). Includes several corrections and amendments. The bibliographies have been extended. An excellent introduction to the history of mathematics. RH

Foundations, P, L.** *History and Philosophy of Modern Mathematics.* Ed: William Aspray, Philip Kitcher. Minnesota Stud. in the Philos. of Sci., V. XI. U of Minnesota Pr, 1988, viii + 386 pp, $35; $13.95 (P). [ISBN: 0-8166-1566-7; 0-8166-1567-5] Proceedings of a May 1985 conference held at the University of Minnesota: fourteen papers by mathematicians (e.g., R. Askey, F. Browder), historians (e.g., H. Edwards, J. Dauben), and philosophers (e.g., M. Friedman, P. Kitcher) giving different perspectives on nineteenth and twentieth century mathematics, prefaced by a lengthy "Opinionated Intro-

ond Edition. David G. Kleinbaum, Lawrence L. Kupper, Keith E. Muller. Stat. & Decision Sci. Prindle, Weber & Schmidt, 1988, xviii + 718 pp. [ISBN: 0-87150-123-6] Regression and related topics along with ANOVA, discriminant analysis, and factor analysis. No calculus or matrices used, but an appendix gives an introduction to matrix algebra. The new edition has some reorganization and new exercises. (*First Edition,* TR, August-September 1979.) FLW

Statistics, P. *Statistical Analysis of Spherical Data.* N.I. Fisher, T. Lewis, B.J.J. Embleton. Cambridge U Pr, 1987, xiv + 329 pp, $65. [ISBN: 0-521-24273-8] A unified treatment of special statistical techniques for the analysis of data in the form of points on a spherical surface or directions in space. Examples illustrate applications in the earth sciences, astrophysics, and other fields. AO

Statistics, P. *Mathematical Statistics and Probability Theory.* Ed: M.L. Puri, *et al.* D Reidel (US Distr: Kluwer Academic), 1987. *Volume A: Theoretical Aspects,* xiii + 326 pp, $68 [ISBN: 90-277-2580-2]; *Volume B: Statistical Inference and Methods,* xiii + 262 pp. [ISBN: 90-277-2581-0] Proceedings of the sixth Pannonian symposium on mathematical statistics held in September 1986 at Bad Tatzmannsdorf, Austria. LAS

Computer Literacy, S(16-17), P, L.** *A Computer Science Reader: Selections from Abacus.* Ed: Eric A. Weiss. Springer-Verlag, 1988, xvi + 447 pp, $35. [ISBN: 0-387-96544-0] A dozen articles (from chess computers to Japanese word processing) and diverse features (editorials, reviews, columns) selected from the first 14 issues of *Abacus,* a popular magazine for computer science professionals akin to *The Mathematical Intelligencer.* A handy compendium of trends and controversies in computer science today. LAS

Elementary Computer Science, T*(13: 1). *Thinking in Pascal: A Systematic Approach.* Daniel Solow. Addison-Wesley, 1988, xxvii + 692 pp, (P). [ISBN: 0-201-12063-1] For the ACM CS1 course. Uses ISO Standard Pascal, with an appendix on Turbo Pascal. Even more emphasis on problem-solving methods and algorithm development than most texts like this, and seems to have many good things to say. Particular attention paid to systematic tracing and debugging techniques. Worth a good look. DFA

Programming, T(13). *An Introduction to Programming with Modula-2.* Patrick D. Terry. Intern. Comput. Sci. Ser. Addison-Wesley, 1987, xii + 460 pp, (P). [ISBN: 0-201-17438-3] For a first course in programming. Chapter 2 (25 pages) introduces algorithms, and the fundamental programming algorithm structures, including case, and loop ...exit; curiously, most exercises in the chapter start with the words "Write a program ...." Syntax diagrams are begun in Chapter 4; I/O is handled through the module EasyInOut (implementation code in Appendix 1). The rest of the book develops usual programming concepts. Includes dynamic data structures. RSF

Programming, T(13). *Modula-2 Discipline & Design.* Arthur Sale. Intern. Comput. Sci. Ser. Addison-Wesley, 1986, xii + 452 pp, (P). [ISBN: 0-201-12921-3] For a first course in programming or a first computer science language course. Chapter 2 introduces lexical elements; Chapter 3 introduces syntax convention, including a two-dimensional notation; the rest of the book develops usual programming concepts. Chapters 9 and 17 are design chapters. InOut is used for I/O. There are chapter reviews and exercises; appendices for syntax and standard module definitions are included. RSF

Languages. *C: The Complete Reference.* Herbert Schildt. Osborne McGraw-Hill, 1987, xv + 773 pp, $24.95 (P). [ISBN: 0-07-881263-1] C is a system's programming language which is widely used to implement such system software as compilers and operating systems. This text is a complete and thorough coverage of the syntax and semantics of this programming language. GMS

Languages, T(13). *Software Engineering in C.* Peter A. Darnell, Philip E. Margolis. Springer-Verlag, 1988, xx + 612 pp, $29.95 (P). [ISBN: 0-387-96574-2] Despite its title, this book is a fairly traditional textbook on C language programming. It describes all features of the C language defined by Kernighan and Ritchie as well as all the features defined in the proposed ANSI C Standard. AO

Computer Systems, P. *Workstations and Publication Systems.* Ed: Rae A. Earnshaw. Springer-Verlag, 1987, xi + 229 pp, $29.50. [ISBN: 0-387-96527-0] The twenty papers in this collection provide an overview of the state-of-the-art in electronic publishing. The papers are organized into four sections covering page description and graphics, document structures and editing, workstations and human-interface aspects, and languages and implementations. AO

Computer Systems, P. *Proceedings of the Eighth TEXUsers Group Annual Meeting.* Ed: Dean Guenther. TEXniques No. 5. TEXUsers Group (POB 9506, Providence, RI 02940), 1988, v + 158 pp, (P). Thirteen papers from the August 1987 annual meeting of the TEXUser's Group featuring such topics as typesetting in Greek, Turkish, and Japanese, device drivers, and diverse applications (journal production, law offices). LAS

Computer Systems, T(18), P. *Concurrency Control and Reliability in Distributed Systems.* Ed: Bharat K. Bhargava. Van Nostrand Reinhold, 1987, xx + 609 pp, $38.95. [ISBN: 0-442-21148-1] Current work on major design principles for building high performance, reliable distributed systems, by 39 contributors. Contains definitions and introductory material for the beginner, theoretical foundations and results, implementation experiences, surveys of many important protocols. DFA

Computer Graphics, P, L. *Image Synthesis: Theory and Practice.* Nadia Magnenat-Thalmann, Daniel Thalmann. Comput. Sci. Workbench.

duction" by the editors. LAS

Linear Algebra, S(14-15), L. *Experiments in Computational Matrix Algebra.* David R. Hill, Cleve B. Moler. Math. Ser. Random House, 1987, xii + 551 pp, $24 (P). [ISBN: 0-394-35678-0] A collection of computational exercises and experiments designed to be carried out using the MATLAB software package. Some discursive material and worked out examples precede each collection of problems. No prior experience with MATLAB is assumed. AO

Algebra, S(17), L. *A Course in Constructive Algebra.* Ray Mines, Fred Richman, Wim Ruitenburg. Universitext. Springer-Verlag, 1988, xi + 344 pp, $32 (P). [ISBN: 0-387-96640-4] Presents the fundamental concepts of modern algebra from a constructive point of view. In the constructive approach to mathematics, the existence of an object can only be proved by establishing a finite routine or algorithm for finding it. Not intended as a first introduction to modern algebra. RH

Algebra, P. *Lecture Notes in Mathematics-1296: Séminaire d'Algèbre Paul Dubreil et Marie-Paule Malliavin.* Ed: M.-P. Malliavin. Springer-Verlag, 1987, iv + 328 pp, $30.30 (P). [ISBN: 0-387-18690-5] A collection of fourteen papers, most of which deal with some aspect of ring theory. SG

Differential Equations, T(15-17: 1). *Second Course in Ordinary Differential Equations for Scientists and Engineers.* Mayer Humi, William Miller. Universitext. Springer-Verlag, 1988, 441 pp, $49 (P). [ISBN: 0-387-96676-5] A book suitable for a second course in differential equations, or a first course in applied mathematics (no partial derivatives) that blends mathematical theory with nontrivial applications from various disciplines. Included in the eleven chapters are Sturm-Liouville theory, special functions (hypergeometric, etc.), linear control theory, Lie groups and algebras, Green's functions, perturbation theory, Liapounov stability, catastrophes and bifurcations, and Sturmian theory. Contains numerous examples, exercises, chapter bibliographies, and an index. RSF

Partial Differential Equations, P. *Mathematical Problems in Viscoelasticity.* Michael Renardy, William J. Hrusa, John A. Nohel. Mono. & Surv. in Pure & Appl. Math., V. 35. Longman Scientific & Technical (US Distr: Wiley), 1987, 273 pp, $87.95. [ISBN: 0-470-20748-5] An overview of results obtained within the past ten years concerning the mathematical analysis of initial and boundary value problems in viscoelasticity—particularly the qualitative behavior of classical solutions. AO

Numerical Analysis, P. *Lectures on Numerical Methods in Bifurcation Problems.* H.B. Keller. Springer-Verlag, 1987, iv + 155 pp, $12 (P). [ISBN: 0-387-18367-1] Local, global, and periodic path-following techniques including methods accounting for singularities and bifurcations. Model problems, theory, applications. RWN

Operator Theory, T(17-18: 1, 2), P, L. *Invertibility and Singularity for Bounded Linear Operators.* Robin Harte. Pure & Appl. Math., V. 109. Marcel Dekker, 1988, xii + 590 pp, $99.75. [ISBN: 0-8247-7754-9] Definition-theorem-proof development of operator theory, culminating in multiparameter spectral theory. Delays the completeness assumption of Banach spaces and discussion of duality until the subject can go no further without them. BC

Analysis, T(16-18: 2), S, P*, L*. *Trigonometric Series, Second Edition, Volumes I & II Combined.* A. Zygmund. Cambridge U Pr, 1988, xiii + 364 pp, $44.50 (P). [ISBN: 0-521-35885-X] Written in 1935 and expanded to two volumes in 1959, this book still stands as a classic in the field of Fourier analysis. A perfect Christmas gift for the budding analyst. (1968 reprint with corrections, TR, August-September 1969.) BC

Differential Geometry, P. *The Mathematics of Surfaces II.* Ed: R.R. Martin. Inst. of Math. & Its Applic. Conf. Ser., V. 11. Clarendon Pr, 1987, xiii + 509 pp, $80. [ISBN: 0-19-853619-4] Papers from the second conference held at the University College, Cardiff, September 1986. Topics focus on techniques for designing new surfaces, fitting them to data, and computer-aided geometric design. MR

Algebraic Topology, T(16-17: 1), S, P, L. *Algebraic Topology via Differential Geometry.* M. Karoubi, C. Leruste. London Math. Soc. Lect. Note Ser., V. 99. Cambridge U Pr, 1987, x + 363 pp, $29.95 (P). [ISBN: 0-521-31714-2] An introduction to algebraic topology with geometry of manifolds, based on De Rham cohomology. This approach has certain inherent advantages, e.g., reduced algebraic formalism, less specialized "mathematical maturity" requirements; drawbacks are a torsionless cohomology, virtually absence of homotopy theory. Translation of a French first course, *sans* exercise sets. RB

Dynamical Systems, T(17-18: 1), P, L. *Dynamical Systems III.* Ed: V.I. Arnold. Ency. of Math. Sci., V. 3. Springer-Verlag, 1988, xiv + 291 pp, $59. [ISBN: 0-387-17002-2] Mathematical aspects of classical and celestial mechanics. Treats the n-body problem, reduction by symmetry, integrable and non-integrable systems, perturbation theory, and small oscillations. BC

Systems Theory, T(17: 1), S, P, L. *State Space and Input-Output Linear Systems.* David F. Delchamps. Springer-Verlag, 1988, x + 425 pp, $48. [ISBN: 0-387-96659-5] A self-contained account beginning with elementary linear algebra, of state-space and input-output systems, including aspects of stability and feedback. BC

Statistics, T(15-18: 1, 2), S*, L*.** *Mathematical Statistics and Data Analysis.* John A. Rice. Stat. & Prob. Ser. Wadsworth, 1988, xx + 594 pp, $32. [ISBN: 0-534-08247-5] Presupposes calculus and linear algebra. Emphasizes data analysis, real problems, and computer use. Data sets available on a floppy disk. Includes a chapter on Bayesian methods. FLW

Statistics, T(15-18: 1, 2), S, L. *Applied Regression Analysis and Other Multivariable Methods, Sec-*

Springer-Verlag, 1987, xv + 400 pp, \$59. [ISBN: 0-387-70023-4] A detailed, up-to-date treatment of both the theoretical and practical aspects of image synthesis. Useful to the non-specialist as an introduction to the subject. AO

Computer Science, P. *The Evolution of Fault-Tolerant Computing.* Ed: A. Avižienis, H. Kopetz, J.C. Laprie. Depend. Comput. & Fault-Tolerant Syst., V. 1. Springer-Verlag, 1987, x + 465 pp, \$85. [ISBN: 0-387-81941-X] The proceedings of a symposium held June 30, 1986 in Baden, Austria in honor of William C. Carter. The eighteen papers in this volume describe the history of work in this area as well as recent research results. AO

Computer Science, P. *Lecture Notes in Computer Science-269: Parallel Algorithms and Architectures.* Ed: A. Albrecht, H. Jung, K. Mehlhorn. Springer-Verlag, 1987, 205 pp, \$21.80 (P). [ISBN: 0-387-18099-0] Proceedings of an international workshop in 1987. Includes invited papers and short communications on models of parallel computation, parallel architectures and hardware algorithms, ánd algorithms for individual problems. RM

Computer Science, S(15-17), P, L. *An Introduction to TCP/IP.* John Davidson. Springer-Verlag, 1988, xii + 100 pp, \$24.95 (P). [ISBN: 0-387-96651-X] Uses the seven-layer ISO reference model to describe the TCP/IP Protocol Suite developed by the Department of Defense for resource sharing computer networks. A good introduction to these protocols. AO

Applications, P. *Artificial Intelligence and Education, Volume One: Learning Environments and Tutoring Systems.* Ed: Robert W. Lawler, Masoud Yazdani. Ablex, 1987, xiv + 439 pp, \$49.50; \$19.95 (P). [ISBN: 0-89391-438-X] Proceedings of the Second International Conference on Artificial Intelligence and Education held in Exeter, United Kingdom, September 1985. Two approaches are presented: Intelligent Tutoring Systems (ITS, Intelligent CIA), focusing on the impact of knowlege engineering concepts on general tools and theory for ITS construction; and computer-based learning environments, for encouraging learning by discovery. RB

Applications (Cognitive Science), P. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition.* David E. Rumehlart, James L. McClelland. MIT Pr, 1986, \$29.95 set (P) [ISBN: 0-262-18123-1]. *Volume 1: Foundations,* xx + 547 pp; *Volume 2: Psychological and Biological Models,* xii + 611 pp. A collection of twenty-six essays describing parallel distributed processing models of the human brain. The models assume that information processing occurs through the interaction of "units"—the basic processing elements—each sending excitatory and inhibatory signals to other units. An appeal of these models is their apparent resemblance to the physiology of the brain. SG

Applications (Communication Theory), T(17: 1), S, P, L. *Signal Detection in Non-Gaussian Noise.* Saleem A. Kassam. Texts in Elect. Engin. Springer-Verlag, 1988, ix + 234 pp, \$48. [ISBN: 0-387-96680-3] Treats three problems in discrete-time signal detection: known, random, and narrowband signals. Noise components here are non-Gaussian, but assumed additive and independent. Exercises with each chapter. BC

Applications (Engineering), T(16: 2), S, L. *Synthetic Aperture Radar.* J. Patrick Fitch. Springer-Verlag, 1988, viii + 170 pp, \$36. [ISBN: 0-387-96665-X] Synthetic Aperture Radar (SAR) is a technique using digital processing to create two-dimensional images from one-dimensional data. It has been used extensively by cartographers and oceanographers since the implementation of the SEASAT and Shuttle Imaging Radar programs. This book provides an excellent introduction to the non-specialist, including good undergraduates, to the techniques of radar imaging. Contains exercises with solutions. MR

Applications (Physics), S(18), P. *The Boltzmann Equation and Its Applications.* Carlo Cercignani. Appl. Math. Ser., V. 67. Springer-Verlag, 1988, xii + 455 pp, \$49.80. [ISBN: 0-387-96637-4] Since its formulation by Boltzmann in 1872 to study the kinetic theory of gases, the Boltzmann equation has found applications in many areas in which only a partial description of a time evolving system is known (e.g., electron transport and radiative transfer). The core chapter of this book establishes existence and uniqueness theorems for solutions to Boltzmann's equation and also describes their qualitative behavior. MR

Applications (Physics), P. *Solvable Models in Quantum Mechanics.* Sergio Albeverio, *et al.* Texts & Mono. in Physics. Springer-Verlag, 1988, xiv + 452 pp, \$79. [ISBN: 0-387-17841-4] The models are given by a Schrödinger Hamiltonian with potential supported on a discrete (perhaps infinite) set of point sources. Cases are considered where the location and strength of the sources are known precisely or given by a probability distribution. MU

Reviewers

RJA: Richard J. Allen, St. Olaf; DFA: David F. Appleyard, Carleton; RB: Richard Brown, Carleton; LB-E: Lisa Brown-Ewing, Macalester; JNC: Judith N. Cederberg, St. Olaf; LC: Laura Chihara, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; RSF: Robert S. Fisk, St. Olaf; SG: Steven Galovich, Carleton; GG: George Gilbert, St. Olaf; BH: Bruce Hanson, St. Olaf; RH: Rhonda Hatcher, St. Olaf; RK: Roger Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; AM: Alan Magnuson, St. Olaf; SM: Steve McKelvey, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; MR: Matthew Richey, St. Olaf; AWR: A. Wayne Roberts, Macalester; GMS: G. Michael Schneider, Macalester; PS: Paul Schoening, St. Olaf; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MS: Myriam Steinback, Macalester; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; LW: Lynne Walling, St. Olaf; FLW: Frank L. Wolf, Carleton; MZ: Michael Zielinski, St. Olaf.

U S Postal Service

STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION

Required by 39 U S C 3685)

1A Title of Publication: THE AMERICAN MATHEMATICAL MONTHLY

1B PUBLICATION NO: 0 0 0 2 9 8 9 0

2 Date of Filing: Sept. 27, 1988

3 Frequency of Issue: monthly except bi-monthly June/July and Aug/Sept

3A No. of Issues Published Annually: ten

3B Annual Subscription Price: Library $90; Indiv. Member $26

4 Complete Mailing Address of Known Office of Publication *(Street City County State and ZIP+4 Code) (Not printers)*
1529 Eighteenth St.,N.W., Washington, D.C. 20036

5 Complete Mailing Address of the Headquarters of General Business Offices of the Publisher *(Not printer)*
1529 Eighteenth St.,N.W., Washington, D.C. 20036

6. Full Names and Complete Mailing Address of Publisher Editor, and Managing Editor *(This item MUST NOT be blank)*

Publisher *(Name and Complete Mailing Address)*
The Mathematical Association of America, 1529 Eighteenth St.,N.W., Washington, D.C. 20036

Editor *(Name and Complete Mailing Address)*
Prof. Herbert S. Wilf, Univ. of Pennsylvania, Dept. of Math.,Philadelphia, PA 19104

Managing Editor *(Name and Complete Mailing Address)*
Harry Waldman, MAA, 1529 Eighteenth St.,N.W., Washington, D.C. 20036

7 Owner *(If owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual must be given. If the publication is published by a nonprofit organization its name and address must be stated.) (Item must be completed.)*

Full Name	Complete Mailing Address
The Mathematical Association of America,Inc.	1529 Eighteenth St.,N.W. Washington, D.C. 20036

8 Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds Mortgages or Other Securities *(If there are none so state)*

Full Name	Complete Mailing Address
None	

9 For Completion by Nonprofit Organizations Authorized To Mail at Special Rates *(DMM Section 423.12 only)*
The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes *(Check one)*

(1) [X] Has Not Changed During Preceding 12 Months
(2) [ ] Has Changed During Preceding 12 Months
(If changed, publisher must submit explanation of change with this statement.)

10 Extent and Nature of Circulation *(See instructions on reverse side)*	Average No. Copies Each Issue During Preceding 12 Months	Actual No. Copies of Single Issue Published Nearest to Filing Date
A Total No Copies *(Net Press Run)*	22,560	20,242
B Paid and/or Requested Circulation 1 Sales through dealers and carriers street vendors and counter sales	0	0
2 Mail Subscription *(Paid and/or requested)*	18,698	19,024
C Total Paid and/or Requested Circulation *(Sum or 10B1 and 10B2)*	18,698	19,024
D Free Distribution by Mail, Carrier or Other Means Samples, Complimentary and Other Free Copies	638	448
E Total Distribution *(Sum of C and D)*	19,336	19,472
F Copies Not Distributed 1. Office use left over unaccounted spoiled after printing	3,224	770
2 Return from News Agents	0	0
G TOTAL *(Sum of E F1 and 2—should equal net press run shown in A)*	22,560	20,242

11 **I certify that the statements made by me above are correct and complete**

Signature and Title of Editor Publisher Business Manager or Owner: Alfred B. Willcox Executive Director, MAA

PS Form 3526, Dec 1987 *(See instructions on reverse)*

INDEX TO VOLUME 95, 1988
THE AMERICAN MATHEMATICAL MONTHLY

TITLE INDEX

AUTHOR INDEX

SUBJECT INDEX

This index uses the 1985 revision of the AMS's 1980 Mathematics Subject Classification.

INDEX OF SPECIAL RECURRING TOPICS

An index of names of proposers and solvers of problems in 1987 and 1988 will appear in this space in 1989.

SOLUTIONS

Numbers in boldface refer to problems; those in lightface to pages.

REVIEWS

Names of authors are in ordinary type; those of reviewers in capitals.

Titles

Authors

THE AMERICAN

MATHEMATICAL MONTHLY

Volume 95, Number 10 December 1988

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Dept. of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Shorter articles about the teaching of mathematics should be sent to either MELVIN HENRIKSEN (linear algebra, algebra, differential equations), Department of Mathematics, Harvey Mudd College, Claremont, CA 91711, or STAN WAGON (calculus, analysis, number theory, discrete mathematics, statistics), Department of Mathematics, Smith College, Northampton, MA 01063. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to ANITA E. SOLOW, Dept. of Mathematics, Grinnell College, Grinnell, IA 50112; in geometry or topology to DENNIS DETURCK, Dept. of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to DAVID J. HALLENBECK, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Proposed problems (three copies) and solutions (two copies), both elementary and advanced to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, University of Calgary, Alberta, Canada T2N 1N4.

Reprint Permission: A. B. WILLCOX, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: MS. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $26 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $29. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $90 per year.

Second class postage paid at Washington, DC, and additional mailing offices.

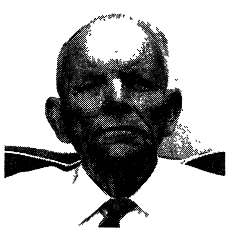
Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Assocation of America, 1529 Eighteenth Street, N.W., Washington, DC 20077-9564.

PRINTED IN THE UNITED STATES OF AMERICA

A New Approach to Bernoulli Polynomials

D. H. LEHMER

Dr. Lehmer needs no introduction to readers of the MONTHLY. It is a pleasure to have a contribution from him in these pages. *Ed.*



Abstract. Beginning with Jacob Bernoulli's discovery before 1705 *of the polynomials that bear his name, there have been five approaches to the theory of Bernoulli polynomials. These can be associated with the names of Bernoulli* [**1**], *Euler* [**2**], *Lucas* [**3**], *Appell* [**4**], *and Hürwitz* [**5**]. *Each mathematician chose to define the Bernoulli polynomials in a different way, and from his definition derived as theorems one or more of the four other definitions. The present article introduces a sixth definition from which the other five are derived.*

Introduction. For the convenience of the reader the first seven Bernoulli polynomials are listed below.

$$
\begin{aligned}
B_0(x) &= 1\\
B_1(x) &= x - \frac{1}{2}\\
B_2(x) &= x^2 - x + \frac{1}{6}\\
B_3(x) &= x^3 - \frac{3}{2}x^2 + \frac{x}{2}\\
B_4(x) &= x^4 - 2x^3 + x^2 - \frac{1}{30}\\
B_5(x) &= x^5 - \frac{5x^4}{2} + \frac{5x^3}{3} - \frac{x}{6}\\
B_6(x) &= x^6 - 3x^5 + \frac{5x^4}{2} - \frac{x^2}{2} + \frac{1}{42}
\end{aligned}
$$

A good account of the formulas involving $B_n(x)$ together with a table of their coefficients up to $n = 15$ will be found in Abramowitz and Stegun [**6**]. Excellent graphs of $B_n(x)$ on the unit interval are given in Davis [**7**].

Five different approaches to the theory of Bernoulli polynomials have been made.

Bernoulli [**1**] (1705?): Sums of powers of the first natural numbers

$$\sum_{k=0}^{m-1} k^{n-1} = \frac{1}{n}\{B_n(m) - B_n(0)\}. \tag{1}$$

Euler [2] (1738): Generating Functions

$$\sum_{n=0}^{\infty} \frac{B_n(x)t^n}{n!} = \frac{te^{xt}}{e^t - 1} \qquad |t| \leqslant 2\pi. \tag{2}$$

Lucas [3] (1891): Umbral Calculus

$$B_n(x) = (B + x)^n. \tag{3}$$

Appell [4] (1832): Appell Sequences

$$B_{n-1}(x) = \frac{1}{n}\frac{d}{dx}B_n(x). \tag{4}$$

Hürwitz [5] (1890): Fourier Series

$$B_n(x) = \frac{-n!}{(2\pi i)^n} \sum_{k=-\infty}^{\infty} k^{-n}e^{2\pi ikx} \qquad (0 < x < 1). \tag{5}$$

In 1851 Raabe [8] discovered the multiplication theorem

$$\frac{1}{m}\sum_{k=0}^{m-1} B_n\left(x + \frac{k}{m}\right) = m^{-n}B_n(mx). \tag{6}$$

It is this identity that we have chosen for a new approach. We begin with two lemmas.

LEMMA 1 (existence). *Let n and m be positive integers. Then there are polynomials of degree n in x satisfying the functional equation*

$$\frac{1}{m}\sum_{k=0}^{m-1} f\left(x + \frac{k}{m}\right) = m^{-n}f(mx). \tag{7}$$

Proof. If $m = 1$, then (7) becomes $f(x) = f(x)$, so that we can assume that $m > 1$. Let us substitute

$$P_n(x) = b_0x^n + b_1x^{n-1} + \cdots + b_n \qquad (b_0 \neq 0) \tag{8}$$

into (7), where the b's are indeterminate coefficients. If f is any solution of (7) so is cf, for any constant c. Hence we may choose $b_0 = 1$.

If we substitute (8) into the left-hand side of (7) we get

$$\frac{1}{m}\sum_{k=0}^{m-1} P_n\left(x + \frac{k}{m}\right) = \sum_{\nu=0}^{n} b_\nu \sum_{\lambda=0}^{n-\nu} x^{n-\nu-\lambda}m^{-\lambda-1}\binom{n-\nu}{\lambda}S_\lambda(m),$$

where

$$S_\lambda(m) = \sum_{k=0}^{m-1} k^\lambda.$$

Collecting the coefficients of x^{n-r} we have

$$\frac{1}{m}\sum_{k=0}^{m-1} P_n\left(x + \frac{k}{m}\right) = \sum_{r=0}^{n} x^{n-r} \sum_{\nu=0}^{n} b_\nu\binom{n-\nu}{r-\nu}m^{\nu-r-1}S_{r-\nu(m)}. \tag{9}$$

Substituting $P_n(x)$ on the right-hand side of (7) we obtain

$$m^{-n}P_n(mx) = \sum_{r=0}^{n} x^{n-r}b_r m^{-r}. \tag{10}$$

Identifying the coefficients of x^{n-r} in (9) and (10) we see that the b's are subject to the condition

$$(m^r - 1)b_r = -\sum_{\nu=0}^{r-1} b_\nu \binom{n-\nu}{r-\nu} m^{\nu-1} S_{r-\nu}(m). \tag{11}$$

In particular when $r = 1$, since $S_1(m) = m(m-1)/2$, we have

$$b_1 = -n/2.$$

In general if

$$b_1, b_2, \dots, b_{r-1}$$

have been determined, (11) serves to determine b_r because $m > 1$. This proves the lemma.

LEMMA 2 (uniqueness). *For a given integer n there is only one monic polynomial of degree n satisfying*

$$\frac{1}{m}\sum_{k=0}^{m-1} f\left(x + \frac{k}{m}\right) = m^{-n}f(mx) \qquad (m > 1). \tag{12}$$

Proof. If possible, let $P_n(x)$ and $Q_n(x)$ be two different monic polynomials of degree n satisfying (12). Suppose that

$$P_n(x) - Q_n(x) = \Delta_d(x) = A_0x^d + A_1x^{d-1} + \cdots$$

where $d < n$ and $A_0 \neq 0$. Subtracting in (12) we get

$$\frac{1}{m}\sum_{k=0}^{m-1} \Delta_d\left(x + \frac{k}{m}\right) = m^{-n}\Delta_d(mx).$$

Identifying the coefficients of x^d on both sides we have

$$A_0 = m^{d-n}A_0,$$

But this contradicts our stipulations that $A_0 \neq 0$, $d < n$ and $m > 1$. This proves the lemma.

We make the following definition:

DEFINITION. The nth Bernoulli polynomial $B_n(x)$ is defined as the unique monic polynomial of degree n which satisfies, for all integers $m > 0$,

$$\frac{1}{m}\sum_{k=0}^{m-1} B_n\left(x + \frac{k}{m}\right) = m^{-n}B_n(mx). \tag{13}$$

THEOREM 1. *The Bernoulli polynomials form an Appell set:*

$$B_{n-1}(x) = \frac{1}{n}\frac{d}{dx}B_n(x) \qquad (n > 0). \tag{14}$$

Proof. If we differentiate the defining relation (13) we get

$$\frac{1}{m}\sum_{k=0}^{m-1} B_n'\left(x+\frac{k}{m}\right) = m^{-n+1}B_n'(mx).$$

This exhibits the monic polynomial $(1/n)B_n'(x)$ as satisfying the functional equation reserved for $B_{n-1}(x)$. The theorem follows from Lemma 2.

THEOREM 2. *Let $B_n = B_n(0)$ be the nth Bernoulli number. Then $B_n(x) = (B+x)^n$, where, after expanding this binomial, the exponents on B are degraded to subscripts.*

Proof. If we continue to differentiate (14) we get

$$\left(\frac{d}{dx}\right)^k B_n(x) = k!\binom{n}{k}B_{n-k}(x). \tag{15}$$

By the Maclaurin expansion of $B_n(x)$

$$\begin{aligned} B_n(x) &= \sum_{k=0}^{n} \frac{x^k}{k!}\left(\frac{d}{dx}\right)^k B_n(0) \\ &= \sum_{k=0}^{n} x^k\binom{n}{k}B_{n-k} = (B+x)^n. \end{aligned}$$

This proves (3).

THEOREM 3. *If $n > 0$*

$$B_n(x+1) - B_n(x) = nx^{n-1}. \tag{16}$$

Proof. The theorem is true for $n = 0$. Suppose the theorem is true for $n = k$ so that

$$B_k(x+1) - B_k(x) = kx^{k-1}. \tag{17}$$

By (14)

$$(k+1)\int_0^x B_k(t)\,dt = B_{k+1}(x) - B_{k+1}(0).$$

Integrating (17) we get

$$\frac{1}{k+1}\big(B_{k+1}(x+1) - B_{k+1}(x)\big) = k\int_0^x t^{k-1}\,dt = x^k.$$

So the theorem is true for $n = k+1$ and, therefore, it is true in general. The identity (1) now follows by summing both sides of (16) over the integers x from 0 to $m-1$.

THEOREM 4. *If $n > 0$,*

$$B_n(x) = \frac{-n!}{(2\pi i)^n}\sum_{r=-\infty}^{\infty}\frac{e^{2\pi irx}}{r^n} \qquad (r \neq 0,\quad 0 < x < 1). \tag{18}$$

Proof. For $n = 1$ the theorem becomes

$$x - \frac{1}{2} = \frac{-1}{2\pi i}\sum_{r\neq 0}\frac{1}{r}e^{2\pi irx} = \frac{-1}{\pi}\sum_{n=1}^{\infty}\frac{\sin 2\pi nx}{n}.$$

This is a well-known fact about Fourier series whose proof is sketched as follows: Let $2\pi x = t + \pi$ so that $x - 1/2 = t/2\pi$. When x ranges from 0 to 1 then t ranges from $-\pi$ to π. Then the identity

$$f(t) = a_0 + \sum_{n=1}^{\infty} (a_n \cos_{nt} + b_n \sin nt),$$

where

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos nt \, dt$$

and

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin nt \, dt$$

becomes when $f(t) = t/2\pi$

$$a_n = \frac{1}{2\pi^2} \int_{-\pi}^{\pi} t \cos nt \, dt = 0$$

$$b_n = \frac{1}{2\pi^2} \int_{-\pi}^{\pi} t \sin nt \, dt = (-1)^{n-1}/n\pi.$$

Thus,

$$x - \frac{1}{2} = \frac{t}{2\pi} = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n\pi} \sin nt = \frac{-1}{\pi} \sum_{n=1}^{\infty} \frac{\sin 2\pi n x}{n}.$$

If we denote the right-hand side of (18) by $\phi_n(x)$, then

$$\frac{1}{n} \frac{d}{dx} \phi_n(x) = \phi_{n-1}(x).$$

Therefore, since

$$\frac{1}{n!} \frac{d^{n-1}}{dx^{n-1}} \phi_n(x) = \phi_1(x) = x - \frac{1}{2},$$

$\phi_n(x)$ is a monic polynomial of the nth degree. Now

$$\frac{1}{m} \sum_{k=0}^{m-1} \phi_n\left(x + \frac{k}{m}\right) = \frac{-n!}{m(2\pi i)^n} \sum_{r \neq 0} \frac{e^{2\pi i r x}}{r^n} \sum_{k=0}^{m-1} e^{2\pi i r k/rn}.$$

The inner sum is m when r is a multiple of m and 0 otherwise. Thus

$$\frac{1}{m} \sum_{k=0}^{m-1} \phi_n\left(x + \frac{k}{m}\right) = \frac{-n!}{m(2\pi i)^n} \sum_{h \neq 0} \frac{1}{(mh)} n e^{2\pi i x m h}$$

$$= m^{-n} \phi_n(mx).$$

By the uniqueness Lemma 2

$$\phi_n(x) = B_n(x).$$

This proves (5) and Theorem 4. Finally we prove (2).

THEOREM 5.

$$\sum_{n=0}^{\infty} B_n(x)t^n/n! = te^{xt}/(e^t - 1).$$

Proof. Let us expand the function

$$F(x, t) = te^{xt}/(e^t - 1)$$

in powers of x and t and collect the coefficients of $t^n/n!$ as a polynomial—$\Psi_n(x)$ of degree n in x:

$$F(x, t) = \sum_{n=0}^{\infty} \Psi_n(x)t^n/n!. \tag{19}$$

Suppose

$$\Psi_n(x) = A_0^{(n)}x^n + A_1^{(n)}x^{n-1} + \cdots + A_n^{(n)}.$$

If we replace x by $1/y$ and t by ty in (19) we get

$$F(1/y, ty) = tye^t/(e^{ty} - 1) = \sum_{n=0}^{\infty} y^n\Psi_n(1/y)t^n/n!.$$

Letting y tend to zero we obtain

$$e^t = \sum_{n=0}^{\infty} A_0^{(n)}t^n/n!$$

Hence, $A_0^{(n)} = 1$ and, hence, $\Psi_n(x)$ is monic.

If in (19) we replace x by $x + k/m$ and sum over k and divide the result by m we get

$$\begin{aligned} \frac{1}{m}\sum_{k=0}^{m-1}\sum_{n=0}^{\infty} \Psi_n(x + k/m)t^n/n! &= \frac{1}{m}\sum_{k=0}^{m-1} F(x + k/m, t) \\ &= \frac{1}{m}\frac{te^{xt}}{e^t - 1}\sum_{k=0}^{m-1} e^{kt/m} = \frac{(t/m)e^{xt}}{e^{t/m} - 1}. \end{aligned} \tag{20}$$

If, instead, we replace in (19) x by mx and t by t/m we obtain

$$F(mx, t/m) = \sum_{n=0}^{\infty} \frac{1}{m}n\Psi_n(mx)t^n/n! = (t/m)e^{xt}/(e^{t/m} - 1). \tag{21}$$

Identifying coefficients of $t^n/n!$ in (21) and (20) we conclude that $\Psi_n(x)$ satisfies the functional equation

$$\frac{1}{m}\sum_{k=0}^{m-1} \Psi_n\left(x + \frac{k}{m}\right) = m^{-n}\Psi_n(mx).$$

Since $\Psi_n(x)$ is monic we have $\Psi_n(x) = B_n(x)$ and (2) is proved.

There are more identities involving Bernoulli polynomials which may be derived from the above. For example, the integral equation

$$\int_0^1 B_n(x + t)\,dt = x^n$$

follows immediately from (6) by letting m tend to infinity.

The functional equation

$$B_n(1-x) = (-1)^n B_n(x)$$

is a simple consequence of (5).

REFERENCES

1. Jacob Bernoulli, Ars conjectandi... Basel 1713, p. 97. Posthumously published. Bernoulli died in 1705.
2. L. Euler, Methodus generalis summandi progressiones, *Comment. acad. sci. Petrop.*, v.6 (1738) 68–97.
3. E. Lucas, Théorie des Nombres, Paris 1891, Chapter 14.
4. P. E. Appell, Sur une classe de polynomes, *Annales d' ecole normal superieur*, s.2, v.9 (1882) 119–144.
5. A. Hürwitz, Personal communication via George Pólya that Hürwitz used the Fourier series approach to Bernoulli polynomials in his lectures.
6. M. Abramowitz and I. Stegun, Editors, Handbook of Mathematical Functions, National Bureau of Standards, Applied Math. Ser. 55, Washington 1964, pp. 804–819.
7. H. T. Davis, Tables of the Higher Mathematical Functions, v.2 1935, Bloomington, p. 185.
8. J. L. Raabe, Zurückführung einiger Summen und bestimmten Integrale auf die Jacob Bernoullische Function, *Journal für die reine and angew. math.*, 42 (1851) 348–376, especially p. 356.

***Addenda, errata, etcetera* for 1988**

1. The Note by Yoram Sagher, *What Pythagoras knew*, in the February, 1988 issue, has drawn several comments to the effect that the points made in the Note had been made before. Professor Nathan J. Fine, in particular, called our attention to his earlier Note in *Math. Mag.* 49, Nov.-Dec. 1976, 249, where virtually the same observations were made. Dr. Fine stated there that Dedekind had already published the idea in 1901.

2. Concerning his article *Angle trisection, the heptagon, and the triskaidecagon*, March, 1988, pp. 185–194, Dr. Andrew Gleason advises us that the following references, all authored by T. I. Zamfirescu, should be mentioned:

(a) Geometric constructions with ruler, compass, and trisector (Romanian), *Stud. Cerc. Mat.*, 15 (1964) 405–411.
(b) Constructibility with ruler, compass, and trisector (Romanian), *Gaz. Mat. Ser. A*, 70 (1965) 204–213.
(c) Constructions with ruler, compass, and trisector (Romanian), *Gaz. Mat. Ser. A*, 71 (1966) 9–18.

The first of the above papers contains a proof of the theorem that describes the integers n for which a regular n-gon can be constructed, using ruler, compass, and trisector.

3. Dr. Doug Hensley, of Texas A & M, has noted an error in *On continued fractions and a certain example of a sequence of continuous functions*, by J. Fabrykowski, on page 539 of the June-July, 1988 issue. Dr. Hensley observes that if $0 < x < 1$ is irrational then x has convergents. These are rationals in $[0,1]$, so they must occur among the r_j. For this infinite sequence of values of j one cannot assert (as claimed in the paper) that $|x - r_j| > 1/(2v_j^2)$. The conclusion is true, nonetheless.

4. Dr. Paul Kumpel, of SUNY Stony Brook, points out that R. Dybvik's letter (95, p. 533) should read '*QFPD* is a rhombus...' instead of '*QFPD* is a parallelogram... .'

Finding Saddlepoints of Two-Person, Zero Sum Games

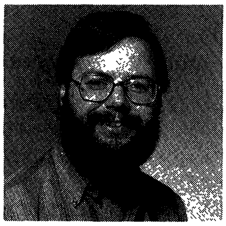
DONNA CRYSTAL LLEWELLYN[1] AND CRAIG TOVEY[2],
School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta

MICHAEL TRICK[3], *Carnegie-Mellon University*

DONNA CRYSTAL LLEWELLYN: During my undergraduate years at Swarthmore College, I became interested in combinatorics and graph theory and their applications. Following these interests, I went to Stanford University and then Cornell University, receiving my doctorate in operations research in 1984 under the direction of L. E. Trotter, Jr. I had the honor of being awarded a National Science Foundation Mathematical Sciences Postdoctorate Fellowship, which I used to pursue my research interests at the University of Bonn, West Germany. Currently, I am an assistant professor at Georgia Institute of Technology in the Department of Industrial and Systems Engineering.

CRAIG TOVEY: As an undergraduate in applied mathematics at Harvard College, I wrote a thesis under the direction of Christos Papadimitriou. Since then my principal research interests have remained in the area of design and analysis of algorithms, particularly probabilistic analysis of combinatorial algorithms and computational complexity. I received my Ph.D. in operations research from Stanford University in 1981, supervised by George Dantzig. Since then I have taught and done research at Georgia Tech, where I am now an associate professor.

MICHAEL TRICK: I received my Ph.D. in Industrial and Systems Engineering from Georgia Tech in 1987. I spent the 1987–88 academic year at the Institute for Mathematics and Its Applications in Minneapolis as a postdoctoral fellow. My research interests include integer programming, design and analysis of algorithms, and applications of operations research in the biological and social sciences. I am now affiliated with the Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh.



1. Introduction. A two-player, zero sum game, as defined by von Neumann and Morgenstern in their classical work **[13]**, is completely specified by its payoff matrix. Let player I have m possible "moves" or "plays" (called *pure strategies*) and player II have n pure strategies. Suppose that player I pays player II a_{ij} if player I chooses his ith strategy and player II chooses her jth. Then the *payoff matrix* of the game is the $m \times n$ matrix $A = [a_{ij}]$.

[1]Supported by NSF Postdoctoral Fellowship DMS84-14104
[2]Supported by NSF Grant ECS-8451032
[3]Supported by ONR Grant N00014-86-K-0173

A *saddlepoint* (SP) of a matrix A is an entry a_{ij}, where a_{ij} is the largest element in row i and the smallest in column j; that is

$$a_{ij} \geqslant a_{ik} \quad \text{for all } k$$
$$a_{ij} \leqslant a_{kj} \quad \text{for all } k.$$

If both inequalities are strict, then we call a_{ij} a *strict saddlepoint* (strict SP).

If there exists a strict saddlepoint of the payoff matrix, then it is the unique optimum solution to the game. A saddlepoint of the payoff matrix (if it exists) corresponds to a (not necessarily unique) optimum solution. Notice that if all of the entries of the payoff matrix are distinct, then a saddlepoint is a strict saddlepoint. (In general, finding the optimal solution vectors for the two players is equivalent to solving an $m \times n$ linear programming problem **[3]**.)

While these properties of saddlepoints are well known, it appears that the question of how to find these solutions efficiently has not been carefully studied. For instance, in **[4]** it is stated that "a [saddlepoint] is easy to find, even in a very large matrix." Other standard texts, such as **[1]**, **[2]**, **[5]**–**[10]**, **[12]** and **[14]**–**[17]** also mention that saddlepoints are easy to locate but fail to discuss finding an efficient algorithm.

The purpose of this article is to show that a strict saddlepoint, when it exists, can be found very quickly. In fact, we give an algorithm that requires looking at only $\simeq n^{0.58}m$ entries of the payoff matrix (where $m \geqslant n$). Hence as the number of strategies increases, *the proportion of the payoff matrix examined approaches zero.* This algorithm also determines when a matrix has no strict SP. Further, we will prove that finding a (nonstrict) saddlepoint requires examination of the entire payoff matrix.

Given a matrix, we refer to the problem of finding a (strict) saddlepoint or determining that one does not exist as the (strict) SP problem. For ease of notation, we will denote $\log_2 x$ by $\lg x$.

We distinguish between two different types of computations. One kind is the ordinary arithmetic operation (i.e., comparison, addition, multiplication and division by integers). The other kind is the identification of the value of a matrix entry; this might involve a lengthy function evaluation. Therefore, we adopt the following model of computation. Assume that each matrix query contributes α units to the running time and each arithmetic operation requires β time; where α and β are some constants.

II. Finding Strict Saddlepoints. We motivate our algorithm for finding a strict saddlepoint with a couple of observations.

Observation 1. If two entries of A, a_{ij} and a_{kl} are known, then at least one of them can be eliminated as a possible strict SP with a single matrix look-up and a constant number of arithmetic operations.

Proof. Clearly this follows if either $i = k$ or $j = l$. Hence assume that neither of these holds. Without loss of generality let $a_{ij} \leqslant a_{kl}$. Look up a_{il}. Clearly, if a_{ij} is a strict SP then $a_{ij} > a_{il}$. Further, if a_{kl} is a strict SP then $a_{il} > a_{kl}$. Since $a_{ij} \leqslant a_{kl}$, it is clear we can not have $a_{ij} > a_{il} > a_{kl}$. Hence, knowing a_{il} will eliminate either a_{ij} or a_{kl} or both from consideration.

Notice that in the proof of observation 1 it is assumed that $a_{ij} \leqslant a_{kl}$. Hence, the entry a_{kj} cannot be a strict SP since $a_{ij} \leqslant a_{kl}$ implies that it is impossible to have $a_{ij} > a_{kj} > a_{kl}$. This leads to our second observation.

Observation 2. If the diagonal terms of a square matrix, A, are known and are such that $a_{11} \leqslant a_{22} \leqslant \cdots \leqslant a_{nn}$, then no term below the diagonal can be a strict SP.

Note that to apply observation 2 we will first have to rearrange the rows and columns of A so that the diagonal is ordered (which will not affect the existence or nonexistence of a strict SP). Our algorithm uses observations 1 and 2 together with proper manipulation of rearrangement pointers. The idea of the algorithm is illustrated in FIGURES 1 and 2.

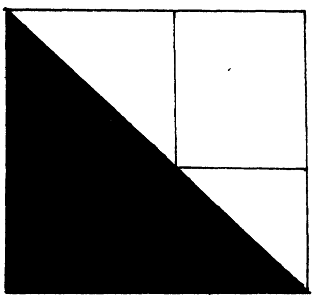


FIG. 1

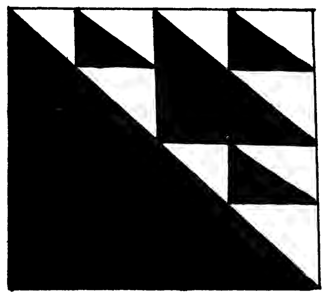


FIG. 2

The main diagonal is looked up and then sorted to eliminate the subdiagonal portion of the matrix from consideration. The remaining region (i.e., the upper triangle) is divided into two triangles and one square which are studied recursively (see FIGURE 1 for the first step of this recursion and FIGURE 2 for the next two steps). These pieces are then put back together using observation 1. In the diagrams, the shaded regions are the parts of the matrix that have been ruled out by our observations and hence do not have to be searched.

We now state the algorithm precisely.

Procedure Square
Input: $A, n, r(1), \ldots, r(n), c(1), \ldots, c(n)$
Output: candidate

This procedure takes as input a matrix A, and a subset of n of its rows and a subset of n of its columns. These rows and columns define an $n \times n$ submatrix $\bar{A}$ of A. The procedure then returns candidate, a location in $\bar{A}$ which might be a strict SP of A. No other element of $\bar{A}$ is a strict SP of A.

Step 0: If $n = 1$, return candidate $= (r(1), c(1))$
Step 1: Look up and sort $A(r(i), c(i))$, $i = 1, \ldots, n$ in ascending order of size.

Suppose σ is the permutation that gives this sorting, i.e., suppose that the set $\{A(r(\sigma(i))), c(\sigma(i))\}_{i=1}^{n}$ is an increasing sequence.

Set

$$r(i) = r(\sigma(i))$$
$$c(i) = c(\sigma(i)) \quad \text{for } i = 1, \ldots, n.$$

Step 2: Call Triangle ($\bar{A}$, n, $r(1), \ldots, r(n), c(1), \ldots, c(n)$; candidate)
Step 3: If candidate $= \varnothing$, return $\varnothing$. Otherwise, take σ^{-1} (candidate) and return it.

Procedure Triangle
Input: $A, n, r(1), \ldots, r(n), c(1), \ldots, c(n)$
Output: candidate

This procedure takes as input a matrix A and a subset of n of its rows and a subset of n of its columns. These rows and columns define an upper triangle submatrix T of A. The procedure then returns a location in T which might be a strict SP of A. No other element of T is a strict SP of A.

Step 0: If $n = 1$, return candidate $= (r(1), c(1))$
Step 1: Call Triangle $(A, \lfloor \frac{n}{2} \rfloor, r(1), \ldots, r(\lfloor \frac{n}{2} \rfloor), c(1), \ldots, c(\lfloor \frac{n}{2} \rfloor)$; candidate1)
Step 2: Call Square $(A, \lceil \frac{n}{2} \rceil, r(1), \ldots, r(\lceil \frac{n}{2} \rceil), c(\lceil \frac{n}{2} \rceil), \ldots, c(n)$; candidate2)
Step 3: Call Triangle $(A, \lfloor \frac{n}{2} \rfloor), r(1 + \lceil \frac{n}{2} \rceil, \ldots, r(n), c(1 + \lceil \frac{n}{2} \rceil), \ldots, c(n)$; candidate3)
Step 4: Using Observation 1, discard at least two of the candidates. If candidate$i$ remains ($i = 1$, 2 or 3), let candidate equal candidate$i$. Otherwise, candidate $= \varnothing$.
Step 5: Return candidate.

Program Strict Saddlepoint
This program solves the strict saddlepoint problem on an $n \times n$ matrix A.
Step 1: Call Square $(n, 1, \ldots, n, 1, \ldots, n$; candidate)
Step 2: If candidate $\neq \varnothing$, check to see if candidate is a strict SP; if so, then report the strict SP. Otherwise, then conclude there does not exist a strict SP.

PROPOSITION. *Program Strict Saddlepoint solves the strict saddlepoint problem for a square matrix.*

Proof. Observation 2 implies that candidate1, candidate2, and candidate3 are the only possible positions of a strict saddlepoint in the upper triangle matrix spanned by the row and column indices which are passed to Procedure Triangle. Let candidate be the output from Procedure Square. Then by the above remark and Step 3 of Procedure Square, this position is the only possible position in the square matrix defined by the indices passed to this procedure by the central program. Thus all that is left for the Program Saddlepoint to do is to check if this position is indeed a strict saddlepoint; which it does in Step 2.

THEOREM. *Program Strict Saddlepoint terminates in time $O(n^{\lg 3})$.*

Proof. Let $F(n)$ be the time required to run Procedure Square $(n, r(1), \ldots, r(n), c(1), \ldots, c(n)$; candidate) and let $G(n)$ be the analogous time for

procedure Triangle $(n, r(1), \ldots, r(n), c(1), \ldots, c(n);$ candidate). Hence, the work required to run Program Saddlepoint is $F(n) + O(\beta n)$; (due to the last check required to see if candidate is really a strict SP in step 2 of Program Saddlepoint). Notice that $G(n) = F(n) - c_1 \beta n \log(\beta n)$ for some constant, c_1, because the only difference in the two algorithms is that the sorting step of Procedure Square does not occur in Procedure Triangle. By analyzing the steps of Program Saddlepoint,

$$F(n) = c_1 \beta n \lg(\beta n) + F\left(\frac{n}{2}\right) + 2G\left(\frac{n}{2}\right) + c_2\beta + c_3\beta n,$$

where the $c_3\beta n$ term is the cost of passing the row and column pointers and $c_2\beta$ is the cost of making the final decision between candidates; that is, step 4 of the triangle procedure (using observation 1).

So,

$$\begin{aligned} F(n) &= c_1\beta n \lg(\beta n) + 3F\left(\frac{n}{2}\right) + c_2\beta + c_3\beta n - 2c_1\beta\left(\frac{n}{2}\right)\lg\left(\frac{\beta n}{2}\right) \\ &= c_1 n \lg n + 3F\left(\frac{n}{2}\right) + c_2\beta + c_3(\beta n) - c_1(\beta n)\lg(\beta n) + c_1(\beta n)\lg\left(\frac{1}{2}\right) \\ &= 3F\left(\frac{n}{2}\right) + c_2\beta + (c_3 - c_1)(\beta n). \end{aligned}$$

Iterating this expression yields

$$F(n) = 3^{\lg n}F(1) + \sum_{i=0}^{\lg n - 1}\left(3^i c_2\beta + \left(\frac{3}{2}\right)^i (c_3 - c_1)(\beta n)\right).$$

But, $F(1) = \alpha$, so after simplifying this geometric series we get

$$\begin{aligned} F(n) &= 3^{\lg n}\left(\alpha + \frac{1}{6}c_2\beta\right) + 2\left(\frac{3}{2}\right)^{\lg n - 1}(c_3 - c_1)(\beta n) - \left(\frac{1}{2}c_2\beta + 2(c_3 - c_1)\beta n\right) \\ &= 3^{\lg n}\left(\alpha + \frac{1}{6}c_2\beta + \frac{4}{3}(c_3 - c_1)\beta\right) - \left(\frac{1}{2}c_2\beta + 2(c_3 - c_1)\beta n\right). \end{aligned}$$

So,

$$F(n) = O(3^{\lg n}).$$

But, $3^{\lg n} = n^{\lg 3}$, and hence

$$F(n) = O(n^{\lg 3}).$$

Hence, by the remark at the beginning of the proof, the Program Strict Saddlepoint runs in time $O(n^{\lg 3})$ as claimed.

Note that this analysis shows that the program saddlepoint requires both order $n^{1.58}$ matrix look-ups and order $n^{1.58}$ total running time. That is, even if a matrix look-up takes significantly more time than an arithmetic operation, that is, $\alpha \gg \beta$, the algorithm will remain $O(n^{1.58})$.

The reader familiar with fractals will note that the unshaded region in FIGURE 2, corresponding to the portion of the matrix that must be searched, is an illustration

of a fractal of dimension lg 3; and hence has area $n^{\lg 3}$. This provides a geometrically intuitive argument for the running time of Program Strict Saddlepoint [**11**, pp. 141–142].

COROLLARY. *The strict SP problem on an $m \times n$ matrix, with $m \geqslant n$, can be solved in $O((\frac{m}{n})n^{\lg 3})$ time.*

Proof. Divide the matrix into $\lceil \frac{m}{n} \rceil$ possibly overlapping $n \times n$ square matrices, $A_1, \dots, A_{\lceil m/n \rceil}$. Apply the theorem to each A_i to get at most $\lceil \frac{m}{n} \rceil$ candidate strict SP's, from which the real strict SP (if it exists) can be found in work $O(\frac{m}{n} + m + n)$. The total work is thus bounded by

$$\left\lceil \frac{m}{n} \right\rceil O(n^{\lg 3}) + O(m) = O\left(\left(\frac{m}{n}\right) n^{\lg 3}\right)$$

as claimed.

III. Saddlepoints. Note that if we are seeking a SP then observation 1 holds only if $a_{11} > a_{22} > \cdots > a_{nn}$ and similarly observation 2 will fail unless the inequality is strict. Thus it is not clear whether the search for a SP can be sped up in a similar way. In fact we show next that the SP problem cannot be solved in time better than cn^2.

Let C be the class of all algorithms that are made up entirely of table look-ups and arithmetic operations.

THEOREM. *There does not exist an algorithm for the SP problem in the class C that can run faster than cn^2. Further, there exists an algorithm in C that does run that fast.*

Proof. We can put a check mark in each matrix cell which is the maximum in its row in time cn^2, and then do the same for the cells which are the minimum in their columns in cn^2 time while checking for a cell with two check marks which would clearly be an SP. Hence, there surely is a quadratic time algorithm in C for this problem.

We now show that no faster algorithm is possible. We do this by playing the adversary against an arbitrary algorithm in C. We argue that the algorithm must look up every entry in A.

Our strategy: For the first $n^2 - 1$ queries of the values of a_{ij} we answer 0, unless all the other entries in row i have been queried, in which case we answer 1. If we answer .5 to the last (n^2) query, then either that cell is the unique SP or there does not exist an SP. This follows since every 0 entry has either a 1 or a .5 in its row and every 1 has either a 0 or a .5 in its column. If instead, we answer -1, then the cell is not a SP, and in fact any zero entry in that last row will be a SP. Hence no algorithm in C can know the answer to the SP problem without n^2 queries.

In conclusion, we have shown that the strict SP problem can be solved in $O(n^{1.58})$ time, but that a similar improvement from $O(n^2)$ for the SP problem cannot be found. Note that our algorithm for the strict SP problem uses $O(n^{1.58})$ matrix queries and has an $O(n^{1.58})$ total running time. We conjecture that it may be possible to solve this problem with less than $O(n^{1.58})$ matrix queries but at the cost of a longer running time.

REFERENCES

1. Jean Pierre Aubin, Mathematical Methods of Game and Economic Theory, North-Holland Publishing Company, Amsterdam, 1979.
2. David Harold Blackwell and Meyer A. Girshick, Theory of Games and Statistical Decisions, Wiley, New York, 1954.
3. George B. Dantzig, Linear Programming and Extensions, Princeton University Press, Princeton, NJ, 1963.
4. Melvin Dresher, L. S. Shapley, and A. W. Tucker, editors, Advances in Game Theory, Princeton University Press, Princeton, NJ, 1964.
5. David Gale, The Theory of Linear Economic Models, McGraw-Hill, New York, 1960.
6. David Gale, The Theory of Matrix Games and Linear Economic Models, Department of Mathematics, Brown University, Providence, RI, 1957.
7. Samuel Karlin, Mathematical Methods and Theory in Games, Programming, and Economics, Addison-Wesley Publishing Company, Reading, MA, 1959.
8. H. W. Kuhn and A. W. Tucker, editors, Contributions to the Theory of Games, Volume I. Princeton University Press, Princeton, NJ, 1950.
9. Richard I. Levin and Robert B. DesJardins, Theory of Games and Strategies, International Textbook Company, Scranton, PA, 1970.
10. Duncan R. Luce and Howard Ráiffa, Games and Decisions: Introduction and Critical Survey, John Wiley and Sons, Inc., New York, 1957.
11. Benoit B. Mandelbrot, The Fractal Geometry of Nature, W. H. Freeman and Company, New York, 1977.
12. Francis B. May, Introduction to Games of Strategy. Allyn and Bacon, Boston, MA, 1970.
13. John von Neumann and Oskar Morgenstern, Theory of Games and Economic Behavior, Princeton University Press, Princeton, NJ, 1944.
14. Guillermo Owen, Game Theory, second edition, Academic Press, New York, 1982.
15. T. Parthasarathy and T. E. S. Raghavan, Some Topics in Two-Person Games, American Elsevier Publishing Company, New York, 1971.
16. Anatol Rapoport, Two Person Game Theory; the Essential Ideas, University of Michigan Press, Ann Arbor, 1966.
17. S. Vajda, The Theory of Games and Linear Programming, Methuen, London, 1967.

Dehn's Algorithm and the Complexity of Word Problems*

RONALD V. BOOK
Department of Mathematics, University of California, Santa Barbara, CA 93106

1. Given a collection of objects such as strings or graphs or trees or... (data structures) and a system of transformation rules that apply to the objects in that collection, there is the following problem:

Instance: two objects from the collection;
Question: can one of the objects be transformed into the other (or does there exist a third object such that both of the given objects be transformed into the third) in a finite number of steps by applying rules from the system?

This problem is called the "word problem" for the given collection. Thue [**27**], [**28**] described this problem for strings, that is, concatenations of letters from a finite alphabet. In this case the transformation rules are ordered pairs (u, v) of strings and the idea of a transformation is that any occurrence of u as a substring may be replaced by v, and vice versa. Thus, the formal version of Thue's problem is the word problem for finitely presented monoids. Thue was also interested in the word problem for other types of combinatorial objects such as graphs and trees (i.e., for data structures). Dehn [**8**] also described this problem, in this case for finitely presented groups, and the work of Dehn became fundamental for combinatorial group theory [**20**]. Word problems have been studied in many contexts in mathematics and in computer science. In any given setting one would like to have an algorithm to solve the word problem. In general, such an algorithm cannot exist since the word problem for finitely presented monoids is undecidable [**21**], [**25**] as is the word problem for finitely presented groups [**4**], [**23**]. (For a general discussion of the existence and nonexistence of algorithms for such problems, see the book by Davis [**7**].)

In cases where the word problem is decidable, one wishes to know the intrinsic computational complexity of the problem. Frequently, the computational complexity of a problem is measured in terms of the "running time" for some fastest algorithm that solves the problem. The running time is the number of steps in the computation of the algorithm on a given input. By considering all inputs of size n and taking the maximum running time over this set of inputs, one obtains an upper bound as a function of n. This function depends on the precise definition of these variables but its asymptotic behavior is invariant under most reasonable definitions. A problem is considered to be tractable if and only if it can be solved in time

---

Ronald Book attended Grinnell College (B.A., 1958) and Wesleyan University (M.A.T., 1960, and M.A., 1964).

After teaching mathematics and physics in high school and teaching mathematics at undergraduate colleges, he studied theoretical computer science at Harvard University, where he earned his Ph.D. (Applied Mathematics) in 1969. His research is in theoretical computer science, primarily structural complexity theory.

*The preparation of this article was supported in part by the National Science Foundation under grant DCR-8314977.

bounded by some polynomial in the size of the input. (For a discussion of tractability, see the book by Garey and Johnson [**11**].)

In the present paper we restrict attention to a class of finitely presented monoids whose presentations are very well behaved. The property of presentations that is of interest is the property of being "Church-Rosser on the congruence class of the identity," a property that is enjoyed by group presentations to which Dehn's Algorithm applies. We will consider this property from the standpoint of rewriting systems and present a simple result that has as a corollary the recent observation of Domanski and Anshel [**10**] that the word problem for groups with Dehn's Algorithm can be solved in linear-time. First we review some of the facts regarding Dehn's Algorithm.

Dehn [**8**] considered presentations of fundamental groups of closed orientable surfaces and showed the appropriate word problems to be decidable. From the standpoint of rewriting systems, Dehn's strategy may be described in the following way.

(i) Suppose that a presentation of a group G, consisting of a finite set Σ of generators and a finite set $R_0 = \{(r_i, 1) | 1 \leqslant i \leqslant m_R\}$ of relators, has the following property: every freely reduced nontrivial word w that is equal to the identity 1 in G has a factorization xyz where for some i, r_i has a factorization yt with y being longer than t. Then R_0 can be transformed into a finite set R_1 of relations for which $(u, v) \in R_1$ implies that the string u is longer than the string v (by placing (y, t^{-1}) in R_1 if $(yt, 1)$ is in R_0). Thus, R_1 permits xyz to be transformed in one step to $xt^{-1}z$, which is shorter. Moreover, $\langle \Sigma, R_1 \rangle$ presents G so that it is sufficient to solve the word problem modulo the group presentation $\langle \Sigma, R_1 \rangle$.

(ii) Consider "reduction rules." The reduction rules in $\langle \Sigma, R_1 \rangle$ are "u reduces to v" for $(u, v) \in R_1$. In addition, the "free reductions" are allowed: "$\sigma\sigma^{-1}$ reduces to 1" and "$\sigma^{-1}\sigma$ reduces to 1" for $\sigma \in \Sigma$. What is known as "Dehn's Algorithm" is the notion of applying the reduction rules in R_1 and the free reductions until no such rule is applicable. For the appropriate group presentations, it is the case that for any w equal to 1 in G, one can apply the reduction rules in R_1 and the free reductions arbitrarily until no such rule is applicable, and *regardless of the order of application of these rules* the final result is 1. Thus, to determine whether w_1 is equal to w_2 in G, it is sufficient to determine whether $w_1 w_2^{-1}$ reduces to 1 by applying *some* finite sequence of the reduction rules.

In the language of rewriting systems, the presentation $\langle \Sigma; R_1 \cup \{(\sigma\sigma^{-1}, 1), (\sigma^{-1}\sigma, 1) | \sigma \in \Sigma\} \rangle$ is "Church-Rosser on the congruence class of the identity" (this notion will be defined in Section 3). Here it is shown that the fact that the word problem for groups with Dehn's Algorithm can be solved in linear-time is a special case of a more general phenomenon: If T is a finite Thue system and w is a string such that T is Church-Rosser (with respect to length) on the congruence class $[w]$ of $w \pmod T$), then there is a linear-time algorithm to determine for an arbitrary string z whether z is congruent to w; if, in addition, T presents a group, then the word problem for that group is solvable in linear-time. The fact that Dehn's Algorithm applies to a finitely presented group G implies that there is a finite Thue system T that presents G (as a monoid) and has the property that T is Church-Rosser on the congruence class of the identity 1, and the observation of Domanski and Anshel follows immediately.

For a description of Dehn's Algorithm from the point of view of combinatorial group theory, see Lyndon and Schupp [**18**] or Zieschang, Vogt, and Coldewey [**30**], and for a description from the point of view of rewriting systems, see Le Chenadec [**16**].

2. Here we provide formal definitions of Thue systems and related notions.

An *alphabet* Σ is a finite set whose members are called *letters*. The *set of words* over Σ is denoted Σ^*, and 1 denotes the empty word. Thus, Σ^* is the free monoid generated by Σ under the operation of concatenation, with the empty word as identity. If $w \in \Sigma^*$, then the *length* of w, denoted $|w|$, is defined as follows: $|1| = 0$, $|a| = 1$ for $a \in \Sigma$, and $|wa| = |w| + 1$ for $w \in \Sigma^*$ and $a \in \Sigma$. The *identity* of words is written as $\equiv$, and the concatenation of words u, v is written as uv.

A *Thue system* T on Σ is a subset of $\Sigma^* \times \Sigma^*$. The *Thue congruence* $\leftarrow * \rightarrow$ *generated by* T is the transitive, reflexive closure of the relation $\leftrightarrow$ defined as follows: for $(u, v) \in T$ and $x, y \in \Sigma^*$, $xuy \leftrightarrow xvy$ and $xvy \leftrightarrow xuy$. Two strings $w, z \in \Sigma^*$ are *congruent* (mod T) if $x \leftarrow * \rightarrow y$, and the *congruence class of* w is $[w] = \{z \in \Sigma^* | z \leftarrow * \rightarrow w\}$. The *monoid presented by* T is defined as follows: (i) the elements are $[x]$, $x \in \Sigma^*$, (ii) the multiplication is $[x] \cdot [y] = [xy]$, $x, y \in \Sigma^*$, and (iii) the identity is $[1]$. This monoid is denoted by $\Sigma^*/\leftarrow * \rightarrow$. If both Σ and T are finite, then $\Sigma^*/\leftarrow * \rightarrow$ is *finitely presented*.

Our main interest here is the word problem for the monoid $\Sigma^*/\leftarrow * \rightarrow$ for any given Thue system. Given T on Σ, this can be expressed as follows:

Instance: two strings x and y in Σ^*;
Question: are x and y congruent (mod T)? (i.e., $[x] = ?[y]$).

In general, this problem is undecidable [**25**], that is, there exists a Thue system T such that no algorithm will decide the word problem for T. We are interested in the situation where the monoid presented by a Thue systems is, in fact, a group and where it is decidable.

Recall that if there is an algorithm to solve the membership problem for a set, then the set is *recursive*. Thus, the word problem for a Thue system T is decidable if and only if for every string z, the congruence class of z modulo T is a recursive set.

3. It is sometimes the case that one wishes to place an ordering on the strings making up each rule of a Thue system. This gives rise to the notion of a "rewriting system." (Rewriting systems are important for the study of automated deduction. See [**14**] and [**29**].)

A *rewriting system* R on Σ is a subset of $\Sigma^* \times \Sigma^*$. An element $(u, v) \in R$ is a *rewriting rule*. The *reduction relation generated by* R is the transitive reflexive closure $= * \Rightarrow$ of the relation $\Rightarrow$ defined as follows: for $(u, v) \in R$ and $x, y \in \Sigma^*$, $xuy \Rightarrow xvy$. A string w is *irreducible* if there is no y such that $w \Rightarrow y$.

For a rewriting system R, denote by R^{-1} the rewriting system $\{(v, u) | (u, v) \in R\}$. Then $R \cup R^{-1}$ is a Thue system and the Thue congruence generated by $R \cup R^{-1}$ is precisely the transitive reflexive closure of the binary relation $(\Rightarrow^{-1} \cup \Rightarrow)$. By the word problem for R, we mean the word problem for the monoid presented by $R \cup R^{-1}$.

Let R be a rewriting system on Σ. We say that R is *noetherian* if there is *no* infinite chain $x_1 \Rightarrow x_2 \Rightarrow \cdots$.

Notice that if R is finite and noetherian, then for $x \in \Sigma^*$ one can compute an irreducible element $\bar{x}$ such that $x = * \Rightarrow \bar{x}$: If x is not irreducible, then it can be reduced; this process can be iterated only finitely many times since R is noetherian. For a string $z \in \Sigma^*$, R is *Church-Rosser on* $[z]$ if $[z]$ contains at most one irreducible element. If R is Church-Rosser on $[z]$ for every $z \in \Sigma^*$, then R is *Church-Rosser*. If R is both noetherian and Church-Rosser, then R is *complete*.

It is easy to see that if R is a finite rewriting system that is complete, then the word problem for R is decidable. Furthermore, if R is a finite rewriting system that is noetherian and for some string z, R is Church-Rosser on $[z]$, then $[z]$ is recursive so that the word problem for $[z]$ is decidable. From Lemma 2 notice if R is a finite noetherian rewriting system on Σ such that the monoid presented by $R \cup R^{-1}$ is a group and such that there is a string z such that R is Church-Rosser on $[z]$, then the word problem for R is decidable.

Now we consider some special systems. Let T be a Thue system on alphabet Σ such that T has no rules that preserve length, i.e., $(u, v) \in T$ implies that $|u| \neq |v|$. Let R be the rewriting system defined by $R = \{(u, v) | (u, v) \in T \text{ or } (v, u) \in T, \text{ and } |u| > |v|\}$. Thus the notion of reduction is based on decreasing the length of the string by applying a rewriting rule. In this context we will use $- * \rightarrow$ to denote the reduction relation $= * \Rightarrow$. Clearly, R is noetherian. By assuming that $(u, v) \in T$ implies $|u| > |v|$, we refer to T instead of to R. Throughout the rest of this paper we will restrict attention to Thue systems with no length-preserving rules and with the notion of reduction being based on decreasing the length of the string by applying a rewriting rule.

The following lemma is very helpful in this context.

LEMMA 1 [2]. *Let T be a finite Thue system on a finite alphabet Σ. There is a linear-time algorithm that on input a string $w \in \Sigma^*$ will compute an irreducible string z such that $w - * \rightarrow z$.*

Proof. Proceed by scanning w from left to right and applying reductions whenever possible. Each time a reduction is applied one must consider the possibility that a portion of the previously scanned string is a prefix of the left-hand side of a reduction and the corresponding suffix of the left-hand side of that reduction is the next portion to be scanned. Thus, backtracking may be needed but the amount of backtracking necessary after a reduction is at most the length of the longest left-hand side of the rules in T. This strategy can be implemented with two pushdown stores, with the first pushdown store being initially empty and the second initially containing w, where the leftmost symbol of w is on the top of the store. When the computation halts, the second pushdown store is empty and the first pushdown store contains an irreducible string z with the rightmost symbol of z on the top of the store. Clearly, at most $|w|$ such reductions can be applied. Hence, the number of steps in the computation is linear in $|w|$ where the constant of linearity depends on m and the number of rules in T.

COROLLARY. *Let T be a finite Thue system on finite alphabet Σ. Suppose that z is a string such that T is Church-Rosser on $[z]$. Then there is a linear-time algorithm to decide the membership problem in $[z]$, i.e., to decide whether a given string is congruent to z.*

COROLLARY [**2**]. *Let T be a finite Thue system on finite alphabet Σ. Suppose that T is Church-Rosser. Then there is a linear-time algorithm to solve the word problem for T.*

The local property of being Church-Rosser on the congruence class of a given string is generally nontrivial since Otto [**24**] has shown that the following question is undecidable:

Instance: a finite Thue system T and a string z:
Question: is T Church-Rosser on $[z]$?

On the other hand, it is known [**3**] that there is a polynomial time algorithm to solve the global property of a finite Thue system being Church-Rosser.

It is clear that if the restriction that no rules are length preserving is omitted and reduction is interpreted as "not increasing the length," then the Church-Rosser property still guarantees that the word problem is decidable. But now a linear time bound cannot be guaranteed; in fact, there are cases where the word problem is complete for polynomial space.

4. Now we turn to the situation where the monoid presented by a finite Thue system is a group. We are interested in the complexity of the word problem for that group in a special case where the word problem is decidable.

Suppose that T is a finite Thue system on alphabet Σ such that $\Sigma^*/\leftarrow * \rightarrow$ is a group. It may be that the inverses of the elements of Σ are not explicitly given; however, the fact that those inverses exist (since $\Sigma^*/\leftarrow * \rightarrow$ is a group) along with the finiteness of T and Σ allows one to compute those inverses. This can be seen as follows. Assume some effective enumeration of the set Σ^*: $w_0, w_1, \ldots$. For each $\sigma \in \Sigma$ and each i, let $y_{\sigma,i}$ denote the string σw_i. Since T is finite, for every string x there are only finitely many strings z that are reachable from x by just one application of a rewriting rule. For each i one can effectively enumerate the set of strings z such that $y_{\sigma,i} \leftarrow * \rightarrow z$ (say by first enumerating all the strings reachable from y_i by means of only one application of a rewriting rule, then enumerating all the strings reachable from $y_{\sigma,i}$ by means of two applications of rewriting rules, etc.). By "dovetailing" these enumerations, we can begin to compute a list of all strings congruent to $y_{\sigma,0}$, all strings congruent to $y_{\sigma,1}, \ldots$. Each time such a string is generated, we check to see if it is identically 1; if $y_{\sigma,j}$ is a string such that $y_{\sigma,j}(\equiv \sigma w_j)$ is congruent to 1, then w_j is congruent to σ^{-1}, so that this computation can halt. But by hypothesis, $\Sigma^*/\leftarrow * \rightarrow$ is a group, so that σ^{-1} has an inverse. This means that eventually some $y_{\sigma,j}$ and z will be enumerated where $y_{\sigma,j}$ is congruent to z and z is identically 1.

Recalling that Σ is finite, let k be the maximum over all $\sigma \in \Sigma$ of the running times of this procedure on σ. Recall that each $w \in \Sigma^*$ has a unique factorization as a product of elements of Σ, say $w \equiv a_1 \ldots a_n$, so that $a_n^{-1} \ldots a_1^{-1}$ is congruent to w^{-1}. Hence, a string that is congruent to w^{-1} can be computed in time $O(k|w|) = O(|w|)$.

This leads to the following fact.

LEMMA 2. *Let T be a finite Thue system on a finite alphabet Σ. Suppose that $\Sigma^*/\leftarrow * \rightarrow$ is a group. Then there is a linear-time algorithm that on input a string $w \in \Sigma^*$ will compute a string z such that z is congruent to w^{-1}, i.e., $[z] = [w]^{-1}$.*

The reader may note that since $\Sigma^*/\leftarrow * \rightarrow$ is a group, one can assume for each element $\sigma \in \Sigma$ there is a string y such that y is congruent to σ^{-1}. Hence, one knows that Lemma 1 holds simply by knowing that the inverses of the elements of Σ exist, that is, those inverses can be treated as constants in the algorithm.

LEMMA 3. *Let T be a finite Thue system on a finite alphabet Σ such that $\Sigma^*/\leftarrow * \rightarrow$ is a group. Suppose that there is a string z such that $[z]$ is a recursive set, that is, there is an algorithm to decide the membership problem for $[z]$. Then the word problem for $\Sigma^*/\leftarrow * \rightarrow$ is decidable.*

Proof. Given x and y in Σ^*, one wishes to determine whether x and y are congruent, i.e., whether $[x] = [y]$. This is accomplished by computing y^{-1} and then determining whether zxy^{-1} is congruent to z. By hypothesis, this question is decidable. Since $\Sigma^*/\leftarrow * \rightarrow$ is a group, cancellation applies, so that zxy^{-1} is congruent to z if and only if xy^{-1} is congruent to 1 if and only if x is congruent to y.

Lemma 3 shows that if $\Sigma^*/\leftarrow * \rightarrow$ is a group and for *some* $z \in \Sigma^*$, $[z]$ is a recursive set, then for *all* $z \in \Sigma^*$, $[z]$ is a recursive set.

Now we consider the situation described in Lemma 1 when the monoid presented by the Thue system is a group. We have the following result.

THEOREM. *Let T be a finite Thue system on alphabet Σ such that the monoid presented by T is a group. Suppose that z is a string in Σ^* and that T is Church-Rosser on $[z]$. Then there is a linear-time algorithm to solve the word problem for this group.*

Proof. As in the proof of Lemma 3, given x and y, one can compute zxy^{-1} and determine whether this string is congruent to z. By Lemmas 1 and 2 this can be determined in time linear in $|zxy^{-1}|$. But T and z are given, so that in general this can be determined in time $O(|xy^{-1}|) = O(|x| + |y|)$.

As a corollary we have the result of Domanski and Anshel.

COROLLARY **[10]**. *Let G be a finitely presented group such that Dehn's Algorithm solves the word problem for G. Then there is a linear-time algorithm to solve the word problem for G.*

One should note that there are other interesting studies of groups with Dehn's Algorithm from the standpoint of rewriting systems. In particular, Greendlinger **[12]**, **[13]** analyzed Dehn's Algorithm from the standpoint of rewriting systems for free groups. Bücken **[5]** showed that if a finitely presented group G has small cancellation, then there is a finite Thue system that presents G and is Church-Rosser on the congruence class of 1. Le Chenadec **[15]** developed a characterization of small cancellation groups by considering a hybrid of rewriting techniques and geometric tools. To see the flavor of current research in the theory of rewriting systems applied to groups and other algebraic structures, see **[16]**, **[17]**.

The study of groups with Church-Rosser presentations is of considerable interest. Muller and Schupp **[22]** characterized the class of groups with context-free word problems. Diekert **[9]** and Madlener and Otto **[19]** studied these groups from the standpoint of Church-Rosser presentations, while Autebert, Boasson, and Senizergues **[1]** studied the type of context-free languages that characterize the congruence class of the identity for such groups.

Squier and Otto [**26**] describe new results on canonical rewriting systems for monoids. Cannon [**6**] describes the system CAYLEY which is designed to solve certain computational problems about finitely presented groups.

REFERENCES

1. J.-M. Autebert, L. Boasson, and G. Senizergues, Groups and NTS languages, *J. Comput. Syst. Sci.*, to appear.
2. R. Book, Confluent and other types of Thue systems, *J. Assoc. Comput. Mach.*, 29 (1982) 171–182.
3. R. Book and C. O'Dunlaing, Testing for the Church-Rosser property, *Theoret. Comput. Sci.*, 16 (1981) 223–229.
4. W. Boone, The word problem, *Annals Math.*, Series 2, 70 (1959), 207–265.
5. H. Bücken, Reduction systems and small cancellation theory, *Proc. 4th Conference on Automated Deduction*, Austin, 1979, 53–59.
6. J. Cannon, The group system CAYLEY, University of Sidney, 1987, preprint.
7. M. Davis, Computability and Unsolvability, McGraw-Hill, 1958. Reprinted by Dover, 1982.
8. M. Dehn, Uber unendliche diskontinuierliche Gruppen, *Math. Ann.*, 71 (1911) 116–144.
9. V. Diekert, Some remarks on presentations by finite Church-Rosser Thue systems, *Proc. 4th Symp. Theoretical Aspects of Computer Science*, Lecture Notes in Computer Science, 247 (1987) Springer-Verlag, 272–285.
10. B. Domanski and M. Anshel, The complexity of Dehn's algorithm for word problems in groups, *J. Algorithms*, 6 (1985) 543–549.
11. M. Garey and D. Johnson, Computers and Intractability, Freeman, 1979.
12. M. Greendlinger, On Dehn's algorithm for the word problem, *Comm. Pure Appl. Math.*, 13 (1960) 67–83.
13. ______, On Dehn's algorithm for the word and continuancy problems, *Comm. Pure. Appl. Math.*, 13 (1960) 641–677.
14. J. P. Jouannaud (ed.). Proc. 1st Int. Conf. Rewriting Techniques and Applications, Lecture Notes on Computer Science, 202 (1985) Springer-Verlag.
15. P. Le Chenadec, Analysis of Dehn's algorithm by critical pairs, Report de Recherche, INRIA, 1985.
16. ______, Canonical Forms in Finitely Presented Algebras, Research Notes in Theoretical Computer Science, Pitman/John Wiley, 1986.
17. ______, A catalogue of complete group presentation, *J. Symb. Computation*, 2 (1986) 363–381.
18. R. Lyndon and P. Schupp, Combinatorial Group Theory, Springer-Verlag, 1977.
19. K. Madlener and F. Otto, Groups presented by certain classes of finite length-reducing string rewriting systems, Proc. 2nd Int. Conf. Rewriting Techniques and Applications, Lecture Notes in Computer Science, 256 (1987) Springer-Verlag, 133–44.
20. W. Magnus, A. Karrass, and D. Solitar, Combinatorial Group Theory, Wiley, 1966. Reprinted by Dover, 1980.
21. A. A. Markov, On the impossibility of certain algorithms in the theory of associative systems, *Doklady Akad. Naut S.S.S.R.*, n.s., 55 (1947), 587–590.
22. D. Muller and P. Schupp, Groups, the theory of ends, and context-free languages, *J. Comput. Syst. Sci.*, 26 (1983) 295–310.
23. P. S. Novikov, On the algorithmic unsolvability of the word problem in group theory, *Trudy Mat. Inst. Steklolv*, 44 (1955) 143.
24. F. Otto, On deciding the confluence of a finite string-rewriting system on a given congruence class, *J.'Comput. Syst. Sci.*, 35 (1987) 285–310.
25. E. Post, Recursive unsolvability of a problem of Thue, *J. Symb. Logic*, 12 (1947) 1–11.
26. C. Squier and F. Otto, The word problem for finitely presented monoids and finite canonical rewriting systems, Proc. 2nd Int. Conf. Rewriting Techniques and Applications, Lecture Notes in Computer Science, 256 (1987) Springer-Verlag, 74–82.
27. A. Thue, Die Lösung eines Spezialfalles eines generellen logischen Problems, K.V.S.S. No. 8, 1910.
28. ______, Probleme über Veränderungen von Zeichenreihen nach gegebenen Regeln, K.V.S.S. No. 10, 1914.
29. L. Wos, Automated reasoning, *Amer. Math. Monthly*, 92 (1985) 85–92.
30. H. Zieschang, E. Vogt, and H.-D. Coldewey, Surfaces and Planar Discontinuous Groups, Lecture Notes in Mathematics 835, 1980, Springer-Verlag.

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

Four Problems on Prime Power Divisibility

A. GARDINER
Department of Mathematics, University of Birmingham, Birmingham B15 2TT, UK

We all know about binomial coefficients, about the sums Σr^{-1} and Σr^{-2}, and about congruences modulo p^r, where p is a prime. Yet combining these simple ingredients, and stirring in a few elementary calculations can still have some surprising results. Our first three problems deal with these very familiar objects.

(A) For which primes p, if any, does the congruence $\binom{2p}{p} \equiv 2$ hold modulo p^4?
(B) For which primes p, if any, does the congruence $\Sigma_1^{p-1} r^{-1} \equiv 0$ hold modulo p^3?
(C) For which primes p, if any, does the congruence $\Sigma_1^{p-1} r^{-2} \equiv 0$ hold modulo p^2?

Elementary calculations suffice to establish that each of these congruences holds modulo any smaller power of p (§§ 2, 3), and that the three questions stated are equivalent (§ 4). Moreover, we shall show that questions (A)–(C) are equivalent to a fourth question (D) (§ 5) which has an interesting history of its own.

(D) For which primes p, if any, does p divide the numerator of the Bernoulli number B_{p-3}?

At present only one such prime is known (see § 1).

1. Irregular primes and Fermat's Last Theorem. We begin by outlining some background to question (D). In 1850 Kummer proved that Fermat's Last Theorem is true for the prime exponent p provided that p is *regular*; that is, provided that p does not divide the numerator of any of the Bernoulli numbers $B_2, B_4, B_6, \ldots, B_{p-3}$. This focussed attention on the *irregular* primes p, which do divide the numerator of one or more of these Bernoulli numbers.

Is there anything special about irregular primes p which divide the numerator of B_{p-3}? Selfridge and Pollack [**5**] listed the irregular primes $p < 25000$ and verified Fermat's Last Theorem in each case. Johnson [**4**] extended (and corrected) this work to cover primes $p < 30000$. He made the explicit observation that, while almost 40% (1273) of the primes p in this range are irregular, only one—namely $p =$ 16843—divides the numerator of the *last* relevant Bernoulli number B_{p-3}. Samuel S. Wagstaff of Purdue University has recently extended this analysis to $p < 150000$, but $p = 16843$ remains the only known irregular prime with this property. Since our

four questions (A)–(D) will be shown to be equivalent, $p = 16843$ is also the only prime < 150000 which satisfies the congruences in (A)–(C).

2. Congruence modulo p. We show first that the congruences in (A), (B), (C) hold modulo p.

(A) We give two proofs.

(i) The binomial coefficient $\binom{2p}{p}$ counts the number of p-element subsets of the set $\{1, 2, \ldots, 2p\}$ of size $2p$. The permutation

$$(1\,2 \cdots p)((p+1)(p+2) \cdots 2p)$$

leaves exactly two of these p-subsets invariant, namely $\{1, 2, \ldots, p\}$ and $\{p+1, p+2, \ldots, 2p\}$, and cycles all the other p-subsets in orbits of length p. Hence

$$\binom{2p}{p} = 2 + (\text{multiple of } p) \equiv 2 \pmod{p}.$$

(ii) The same congruence modulo p also follows from the algebraic definition of the binomial coefficient

$$\binom{2p}{p} = \frac{(2p)!}{p!p!} = 2 \cdot \frac{(2p-1)(2p-2) \cdots (p+1)}{(p-1)!}.$$

Since the denominator $(p-1)!$ is not a multiple of p, it is invertible modulo p. Moreover $(2p-1)(2p-2) \cdots (p+1) \equiv (p-1)! \pmod{p}$, so

$$\binom{2p}{p} \equiv 2 \cdot (p-1)!((p-1)!)^{-1} \equiv 2 \pmod{p}. \qquad \blacksquare$$

(B) The congruence in (B) is false modulo 2, so let p be an odd prime. Distinct non-zero residues modulo p have distinct inverses modulo p. Thus as r runs through the residues from 1 to $p-1$ modulo p, so also does r^{-1}. Hence

$$\sum_{1}^{p-1} r^{-1} \equiv \sum_{1}^{p-1} i \equiv \sum_{1}^{(p-1)/2} (i + (p-i)) \equiv 0 \pmod{p}. \qquad \blacksquare$$

(C) The congruence in (C) is false both modulo 2 and modulo 3, so let p be a prime $\geqslant 5$. As r runs through the nonzero residues, so also does r^{-1}. Hence

$$\sum_{1}^{p-1} r^{-2} \equiv \sum_{1}^{p-1} (r^{-1})^2 \equiv \sum_{1}^{p-1} i^2 \pmod{p}.$$

Since 6 is invertible modulo p when $p \geqslant 5$, the familiar formula

$$\sum_{1}^{p-1} i^2 = \frac{1}{6} p(p-1)(2p-1)$$

yields

$$\sum_{1}^{p-1} r^{-2} \equiv 0 \pmod{p}.$$

3. Congruence modulo p^r ($r \geqslant 2$).

(B) We begin by showing that the congruence in (B) holds modulo p^2 provided $p \geqslant 5$ [2]. (The congruence is clearly false modulo 3^2.) The identity

$$\sum_1^{p-1} \frac{1}{r} = \sum_1^{(p-1)/2} \left(\frac{1}{r} + \frac{1}{p-r} \right) = p \cdot \sum_1^{(p-1)/2} \frac{1}{r(p-r)} = \frac{1}{2} p \cdot \sum_1^{p-1} \frac{1}{r(p-r)}$$

works just as well modulo p^2, so

$$\begin{aligned} \sum_1^{p-1} r^{-1} &\equiv 2^{-1} \cdot p \cdot \sum_1^{p-1} r^{-1}(p-r)^{-1} \pmod{p^2} \\ &\equiv 2^{-1} \cdot p \cdot \left(- \sum_1^{p-1} r^{-2} \right) \pmod{p^2} \\ &\equiv 0 \pmod{p^2}. \end{aligned}$$

■

(A) The congruence in (A) holds modulo p^2 when $p = 3$, and modulo p^3 when $p \geqslant 5$. (Dickson attributes this to Wolstenholme (1862) and to Cunningham (1907) [1].) This is the special case "$s = 0$, $k = 2$" of the following theorem.

THEOREM 1. *Let p be any prime $\geqslant 5$. If the positive integer k is a multiple of p^s, then*

$$\binom{kp}{p} \equiv k \pmod{p^{2s+3}}.$$

Proof. Let $k = mp^s$, and put $t = s + 1$. Then $kp = mp^t$. Partition a set of size mp^t into m parts of size p^t. Then the number of p-subsets of the whole set satisfies

$$\binom{mp^t}{p} = \sum_{i_1 + i_2 + \cdots + i_m = p} \binom{p^t}{i_1}\binom{p^t}{i_2} \cdots \binom{p^t}{i_m}.$$

The m terms with exactly one $i_j \neq 0$ are all equal to $\binom{p^t}{p}$. The terms with three or more of the $i_j \neq 0$ are all divisible by p^{3t}. Hence

$$\binom{mp^t}{p} \equiv m\binom{p^t}{p} + \sum_1^{p-1} \binom{p^t}{i}\binom{p^t}{p-i} \pmod{p^{3t}}. \tag{1}$$

We consider the two terms on the right-hand side separately:

$$\begin{aligned} \binom{p^t}{p} &= \frac{p^t(p^t - 1) \cdots (p^t - (p-1))}{p!} \\ &= p^s \cdot \frac{1}{(p-1)!} \left((p-1)! - p^t \cdot \sum_1^{p-1} \frac{(p-1)!}{r} + \left(\text{multiple of } p^{2t}\right) \right). \end{aligned}$$

If we interpret this identity modulo p^{2s+3} and use the fact that, since $p \geqslant 5$, we know that the congruence in (B) holds modulo p^2, then we obtain

$$\binom{p^t}{p} \equiv p^s \pmod{p^{2s+3}}. \tag{2}$$

Expanding the summands in the second term of (1) in the same way we get

$$\binom{p^t}{i}\binom{p^t}{p-i} = \frac{p^t(p^t-1)\cdots(p^t-(i-1))}{i!} \cdot \frac{p^t(p^t-1)\cdots(p^t-(p-i-1))}{(p-i)!}$$

$$\equiv -p^{2t}\cdot i^{-1}(p-i)^{-1} \pmod{p^{3t}}.$$

$$\sum_1^{p-1}\binom{p^t}{i}\binom{p^t}{p-i} \equiv -p^{2t}\cdot\sum_1^{p-1} i^{-1}(p-i)^{-1} \pmod{p^{3t}}$$

$$\equiv p^{2t}\cdot\sum_1^{p-1} i^{-2} \pmod{p^{2t+1}}$$

$$\equiv 0 \pmod{p^{2t+1}}, \tag{3}$$

since $\sum_1^{p-1} i^{-2} \equiv 0 \pmod p$. Combining (1), (2), and (3) gives the result.

4. The unsolved problems (A)–(C). We show first that (A) and (C) are equivalent. We then show that (B) and (C) are equivalent.

The fact that (A) and (C) are equivalent follows from a natural attempt to extend Theorem 1. When does the congruence in Theorem 1 hold modulo a higher power of the prime p? It certainly does if $k \equiv 0 \pmod{p^{s+1}}$. But it can also happen in other ways: for example $\binom{6.5}{5} = 142506 \equiv 6 \pmod{5^4}$. This suggests the question

(A′) For which primes p, and which integers $m \not\equiv 0 \pmod p$ does the congruence $\binom{mp^{s+1}}{p} \equiv m$ hold modulo p^{2s+4}?

When $s = 0$ a straightforward expansion of the binomial coefficient gives

THEOREM 2. *Let p be any prime $\geqslant 5$. If m is not a multiple of p, then* $\binom{mp}{p} \equiv m \pmod{p^4}$ *holds if and only if either* (a) $m \equiv 1 \pmod p$, *or* (b) $\sum_1^{p-1} r^{-2} \equiv 0 \pmod{p^2}$.

COROLLARY. *Questions* (A) *and* (C) *have the same answer.*

We now ask when the congruence in (B) holds modulo p^3. The calculation used in § 2 to prove this congruence modulo p^2 already contained a strong hint that this is closely related to question (C).

THEOREM 3. *Let p be any prime $\geqslant 3$. Then* $\sum_1^{p-1} r^{-1} \equiv 0 \pmod{p^3}$ *holds if and only if* $\sum_1^{p-1} r^{-2} \equiv 0 \pmod{p^2}$.

Proof.

$$2p\cdot\sum_1^{p-1} r^{-2} = 2p\cdot\sum_1^{p-1}(r^2)^{-1}$$

$$\equiv -2p\cdot\sum_1^{p-1}(p^2-r^2)^{-1} \pmod{p^3}. \tag{4}$$

When $1 \leqslant r \leqslant p-1$ the identity

$$\frac{2p}{p^2-r^2} = \frac{1}{p-r} + \frac{1}{p+r}$$

has a natural analogue modulo p^3. If we use this, then (4) becomes

$$-2p \cdot \sum_1^{p-1} r^{-2} \equiv \sum_1^{p-1} (p-r)^{-1} + \sum_1^{p-1} (p+r)^{-1} \pmod{p^3}$$

$$\equiv \sum_1^{p-1} r^{-1} + \sum_1^{p-1} (p+r)^{-1} \pmod{p^3}. \tag{5}$$

Since $1 \leqslant r \leqslant p-1$ we have the following alternative expression for $(p+r)^{-1}$

$$(p+r)^{-1} \equiv r^{-1} - pr^{-2} + p^2 r^{-3} \pmod{p^3}.$$

Substituting this into (5) now gives

$$-2p \cdot \sum_1^{p-1} r^{-2} \equiv 2 \cdot \sum_1^{p-1} r^{-1} - p \cdot \sum_1^{p-1} r^{-2} + p^2 \cdot \sum_1^{p-1} r^{-3} \pmod{p^3}. \tag{6}$$

Finally, the formula for the sum of the first $p-1$ cubes shows that

$$\sum_1^{p-1} r^{-3} \equiv \sum_1^{p-1} i^3 \equiv \left(\frac{p(p-1)}{2}\right)^2 \equiv 0 \pmod{p^2}$$

$$-p \cdot \sum_1^{p-1} r^{-2} \equiv 2 \cdot \sum_1^{p-1} r^{-1} \pmod{p^3}.$$

The theorem follows. ■

COROLLARY 1. *Questions* (B) *and* (C) *have the same answer.*

COROLLARY 2. *Let p be any prime $\geqslant 5$. If $m \not\equiv 0, 1 \pmod p$, then the congruences* (a) $\binom{mp}{p} \equiv m \pmod{p^4}$, (b) $\Sigma_1^{p-1} r^{-1} \equiv 0 \pmod{p^3}$, (c) $\Sigma_1^{p-1} r^{-2} \equiv 0 \pmod{p^2}$, *are all equivalent.*

5. Bernoulli numbers. We show that questions (C) and (D) are equivalent. It follows that all four of our questions (A)–(D) have the same answer.

Jacob Bernoulli posed the problem of expressing the sum

$$S_m(n) = 1^m + 2^m + \cdots + (n-1)^m$$

as a polynomial in n. He gave an inductive method for finding $S_m(n)$ for any given m [**3**, § 15.1]. The constant term of this polynomial is -1 when $m = 0$ and zero when $m \geqslant 1$. The coefficient of n in $S_m(n)$ is defined to be B_m, the mth Bernoulli number. Thus $B_0 = 1$, $B_1 = -1/2$, $B_2 = 1/6$, $B_3 = 0$, and so on.

Let $B_m = U_m/V_m$ where U_m and V_m are relatively prime integers. If p is a prime $\geqslant 5$, then $S_{p-3}(p) \equiv 0 \pmod p$ [**3**, p. 235], and $V_{p-3} \cdot S_{p-3}(p) \equiv U_{p-3} \cdot p \pmod{p^2}$ [**3**, Proposition 15.2.2]. Hence p divides the numerator U_{p-3} of B_{p-3} if

and only if $S_{p-3}(p) \equiv 0 \pmod{p^2}$. A simple calculation then shows that (for $p \geqslant 5$)

$$S_{p-3}(p) \equiv \sum_{1}^{p-1} r^{-2} \pmod{p^2}.$$

Hence questions (C) and (D) have the same answer.

Acknowledgement. I am grateful to D. R. Heath-Brown for directing my attention to question (D).

REFERENCES

1. L. E. Dickson, History of the Theory of Numbers, Vol. 1, Chelsea, New York, 1952, pp. 271, 274.
2. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, 4th ed., Oxford University Press, 1960, p. 89.
3. K. Ireland and M. Rosen, A Classical Introduction to Modern Number Theory, Springer, New York, 1982.
4. Wells Johnson, Irregular primes and cyclotomic invariants, *Math. Comp.*, 29 (1975) 113–120.
5. J. L. Selfridge and B. W. Pollack, Fermat's last theorem is true for any exponent up to 25,000, *Notices Amer. Math. Soc.*, 11 (1964) 97.

NOTES

EDITED BY DAVID J. HALLENBECK, DENNIS DETURCK, AND ANITA E. SOLOW

Another Brief Proof of the Sylvester Theorem

X. B. LIN
Mathematics Department, Michigan State University, East Lansing, MI 48824

A finite set of points, S, in a projective or affine space such that no lines intersects S in exactly two points is known as a Sylvester-Gallai (SG) configuration. In real space it is well known that there are no nonlinear SG's. There are several simple proofs of this fact (see **[1]**–**[4]**) and we offer here still another, as far as we know somewhat different from the others. The result in E^n follows from that in E^2 by projection.

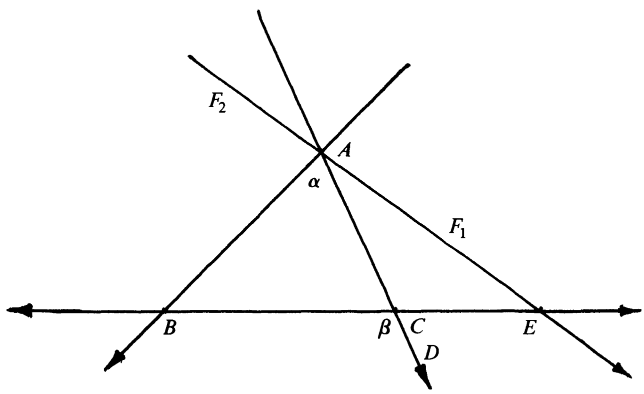
THEOREM. *If S is a finite set of points in E^2 such that no line intersects S in precisely two points, the S is a subset of a line.*

Proof. We use Hilbert's definition of angle as the union of two noncollinear rays with a common end point. If A, B, and C are in S, then the angle defined by rays $\overrightarrow{AB}$ and $\overrightarrow{AC}$ is denoted BAC or CAB. If $\overrightarrow{AB} \cup \overrightarrow{AC}$ contains a fourth point of S, the angle is called an admissible angle (relative to S). Suppose now that S is not linear. Then the angles at the vertices of the convex hull of S are certainly admissible so the set of such angles is not empty. Suppose BAC is the largest such angle with angle measure $\alpha < 180$ and that D is a fourth point of S on ray $\overrightarrow{AC}$. There is no loss of generality in assuming that C is between A and D.

Now BCD is larger than BAC so it is not an admissible angle. Thus ray $\overrightarrow{CB}$ cannot contain a third point of S. But the line $\overrightarrow{CB}$, by assumption, must contain another point E of S so E must be on the ray opposite to $\overrightarrow{CB}$.

But now the line AE presents us with a contradiction, since it must contain a third point F of S, and if F is on ray $\overrightarrow{AE}$, then BAE is admissible and larger than BAC; while if F is on the ray opposite to $\overrightarrow{AE}$, then FAC is admissible and greater than BAC.

This contradiction shows that S must be linear.



REFERENCES

1. H. S. M. Coxeter, Int. to Geom., 2nd ed., 1969, John Wiley and Sons, N.Y., pp. 65–66.
2. L. M. Kelly and W. O. S. Moser, On the number of ordinary lines determined by *n* points, *Canadian Jour. of Math.*, (10) 1958, 213.
3. Editorial Note, *Amer. Math. Monthly*, 51 (1944) 170–171.
4. G. D. Chakerian, Sylvester's Problem on Collinear Points and a Relative, *Amer. Math. Monthly*, 77 (1970) 164–167.

$A \geqq B \geqq 0$ Ensures $(BA^2B)^{1/2} \geqq B^2$— Solution to a Conjecture on Operator Inequalities

TAKAYUKI FURUTA
Department of Mathematics, Faculty of Science, Hirosaki University, Bunkyo-cho 3, Hirosaki 036, Aomori, Japan

DEDICATED TO PROFESSOR ZIRÔ TAKEDA WITH RESPECT AND AFFECTION

In an issue of this MONTHLY [**1**, p. 539], the following conjecture is stated.

CONJECTURE. *Let A and B be hermitian matrices on a finite dimensional Euclidean space. If* $A \geqq B \geqq 0$, *then*

$$(BA^2B)^{1/2} \geqq B^2$$

and

$$A^2 \geqq (AB^2A)^{1/2}.$$

In this short note, we shall prove this conjecture in a more general form.

THEOREM 1. *If A and B are positive bounded hermitian linear operators on a Hilbert space such that* $A \geqq B \geqq 0$, *then*

$$(BA^2B)^{1/2} \geqq B^2 \tag{1}$$

and

$$A^2 \geqq (AB^2A)^{1/2}. \tag{2}$$

We prove the following Lemma needed for Theorem 1.

LEMMA. *If A and B are positive bounded hermitian linear operators on a Hilbert space such that* $A \geqq B \geqq 0$, *then*

$$(B^{1/2}A^3B^{1/2})^{1/2} \geqq B^2 \tag{i}$$

and

$$(B^{1/2}A^2B^{1/2})^{1/3} \geqq B. \tag{ii}$$

We quote the following result to show the Lemma.

THEOREM A ([**2**]). *Let X and Y be bounded linear operators on a Hilbert space H. We suppose that* $X \geqq 0$ *and* $\|Y\| \leqq 1$. *If f is an operator monotone function defined on*

$[0, \infty[$, *then*

$$f(Y^*XY) \geqq Y^*f(X)Y.$$

Proof of the Lemma. If $A \geqq B \geqq 0$, then $A + \varepsilon \geqq B + \varepsilon \geqq \varepsilon$ for any $\varepsilon > 0$, so $B + \varepsilon$ and $A + \varepsilon$ are both invertible and since for any bounded linear operator T, $(T + \varepsilon)^p$ converges to T^p in norm as $\varepsilon \to 0$ for each $p > 0$, therefore we may assume that A and B are invertible. If $A \geqq B \geqq 0$, then $B^{1/2}A^{-1}B^{1/2} \leqq I$, hence $\|A^{-1/2}B^{1/2}\| \leqq 1$. Put $Y = A^{-1/2}B^{1/2}$. By Theorem A, we have $f(Y^*XY) \geqq Y^*f(X)Y$ for an operator monotone function f defined on $[0, \infty[$ and for $X \geqq 0$ since $\|Y\| \leqq 1$. Take $X = A^4$ and $f(t) = t^{1/2}$ for real $t \geqq 0$, then $f(t)$ is an operator monotone function ([3]) and $X \geqq 0$. By Theorem A we have

$$\begin{aligned}\left(B^{1/2}A^{-1/2}A^4A^{-1/2}B^{1/2}\right)^{1/2} &\geqq B^{1/2}A^{-1/2}A^{4/2}A^{-1/2}B^{1/2}\\ &\geqq B^{1/2}AB^{1/2} \geqq B^{1/2}BB^{1/2} = B^2\end{aligned}$$

that is, (i) of the Lemma is shown. Similarly, take $X = A^3$ and $f(t) = t^{1/3}$, then we have (ii) in the Lemma.

Proof of Theorem 1. Put $C = (B^{1/2}A^2B^{1/2})^{1/3}$. Then $C \geqq B \geqq 0$ by (ii) in the Lemma, so that applying (i) in the Lemma to C and B, we have

$$\left(B^{1/2}C^3B^{1/2}\right)^{1/2} \geqq B^2,$$

that is,

$$\left(B^{1/2}B^{1/2}A^2B^{1/2}B^{1/2}\right)^{1/2} \geqq B^2,$$

whence

$$\left(BA^2B\right)^{1/2} \geqq B^2$$

so that (1) is shown. (2) is an easy consequence of (1) taking inverses. This is shown in [1] and we reproduce it for the sake of convenience. We may assume the existence of B^{-1} without loss of generality. By hypothesis, $B^{-1} \geqq A^{-1} \geqq 0$. (1) in Theorem 1 ensures $(A^{-1}B^{-2}A^{-1})^{1/2} \geqq A^{-2}$. Taking the inverse gives (2).

Remark. Further extensions of Theorem 1 will appear in my forthcoming paper together with some related results.

Addendum. We give an elementary proof of Theorem 1. Since $A \geqq B \geqq 0$ ensures $A^\alpha \geqq B^\alpha$ for any $\alpha \in [0.1]$ ([3]), for any $r \in [0, 1/2]$ we have

$$\left(B^rA^2B^r\right)^{1/2} = \left(B^rA^{1-r}A^{2r}A^{1-r}B^r\right)^{1/2} \geqq \left(B^rA^{1-r}B^{2r}A^{1-r}B^r\right)^{1/2} \geqq B^rA^{1-r}B^r \geqq B^{1+r}. \tag{*}$$

Put $r = 1/2$ in (*), so $C \equiv (B^{1/2}A^2B^{1/2})^{1/2} \geqq B^{3/2} \equiv D \geqq 0$. Then applying (*) to $C \geqq D \geqq 0$ and put $r = 1/3$, we have $(D^{1/3}C^2D^{1/3})^{1/2} \geqq D^{4/3}$, that is, $(BA^2B)^{1/2} \geqq B^2$. We remark that $(B^rA^2B^r)^{1/2} \geqq B^{1+r}$ holds for any $r \geqq 0$ by the same method stated above.

REFERENCES

1. N. N. Chan and Man Kam Kwong, Hermitian matrix inequalities and a conjecture, this MONTHLY, 92 (1985) 533–541.
2. F. Hansen, An operator inequality, *Math. Ann.*, 246 (1980) 249–250.
3. G. K. Pedersen, Some operator monotone function, *Proc. Amer. Math. Soc.*, 36 (1972) 309–310.

The Evaluation of a Putnam Integral

S. J. BERNAU

Department of Mathematical Sciences, The University of Texas, El Paso, TX 79968

One of the problems on the 1985 Putnam Competition [1] was the evaluation of $\int_0^\infty t^{-1/2}e^{-1985(t+t^{-1})}\,dt$, with the value $\sqrt{\pi}$ given for $\int_{-\infty}^\infty e^{-x^2}\,dx$. The suggested solution sets $a = 1985$ and considers

$$I(x) = \int_0^\infty t^{-1/2}e^{-at-xt^{-1}}\,dt.$$

Differentiation under the integral sign, and changes of variable lead to a first order differential equation for $I(x)$. The solution of this gives

$$I(x) = I(0)e^{-2(ax)^{1/2}},$$

with $I(0)$ obtainable from the given integral as $\sqrt{\pi/a}$.

The purpose of this note is to point out a simpler method which does not use the auxiliary variable x, nor the differentiation under the integral sign.

Suppose $a > 0$, and let

$$J(a) = \int_0^\infty t^{-1/2}e^{-a(t+t^{-1})}\,dt.$$

We have

$$J(a) = \int_0^1 t^{-1/2}e^{-a(t+t^{-1})}\,dt + \int_1^\infty t^{-1/2}e^{-a(t+t^{-1})}\,dt.$$

Substitute $1/t$ for t in the first integral to conclude

$$J(a) = \int_1^\infty (t^{-1/2} + t^{-3/2})e^{-a(t+t^{-1})}\,dt.$$

Now make the substitution $u = a^{1/2}(t^{1/2} - t^{-1/2})$ to obtain

$$J(a) = 2a^{-1/2}\int_0^\infty e^{-u^2-2a}\,du = \sqrt{\frac{\pi}{a}}\,e^{-2a}.$$

Our method is also strong enough to recapture the two parameter integral above. In the integral for $I(x)$ set $v = (a/x)^{1/2}t$ to conclude

$$I(x) = (x/a)^{1/4}J\big((ax)^{1/2}\big) = (x/a)^{1/4}\pi^{1/2}(ax)^{-1/2}e^{-2(ax)^{1/2}}$$
$$= \sqrt{\frac{\pi}{a}}\,e^{-2(ax)^{1/2}},$$

as in [1].

REFERENCE

1. Leonard F. Klosinski, G. L. Alexanderson, and Loren C. Larson, The William Lowell Putnam Mathematical Competition, The Amer. Math. Monthly, 93 (1986) 620–626.

A Generalization of Ceva's Theorem to Higher Dimensions

STEVEN LANDY
1702 Arley Dr., Indianapolis, IN 46229

Ceva's theorem in plane geometry gives a necessary and sufficient condition for the concurrence of three cevians in a triangle [**1**, p. 4]. A cevian is defined to be a segment from a vertex to the side opposite it. With reference to FIGURE 1, Ceva's theorem states that AX, BY, and CZ are concurrent in point P if and only if

$$\frac{AZ}{BZ}\cdot\frac{BX}{CX}\cdot\frac{CY}{AY}=1. \tag{1}$$

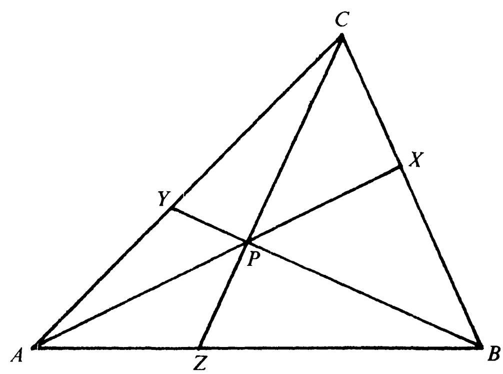


FIG. 1

A generalization of (1) has been found for pyramids in higher dimensional spaces. N points in an $N-1$ dimensional Euclidean space determine an $N-1$ dimensional pyramid. Define a "cevian" as a segment which joins a vertex of that pyramid to a point (hereafter called a "base point") in the $N-2$ dimensional cell opposite it. A general Ceva theorem may then be stated as follows:

N cevians of an $N-1$ dimensional pyramid are concurrent if and only if a vertex mass assignment exists such that each base point is at the center of mass of its surrounding cell.

Proof. In one direction the theorem is obvious. If each vertex may be assigned a mass such that the base points are all cell cm's, then each of the cevians must pass through the cm of all N vertices. So all the cevians will be concurrent at that point. Therefore, a mass assignment is sufficient for a "Ceva configuration."

Now assume a Ceva configuration exists for an $N-1$ dimensional pyramid. Let A and B be two of the vertices and X and Y their opposite base points (see FIGURE 2). AX and BY intersect in P (the "Ceva point") and so determine a plane Q. Q intersects the $N-3$ dimensional pyramid, R, generated by the $N-2$ remaining vertices $C, D, \dots E$. AY and BX extended, being in Q, and being not parallel, intersect in a point F lying in R. We now recall that the vector to any point within

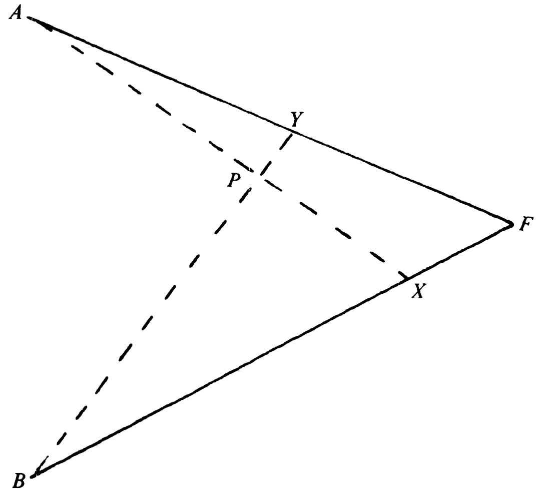


FIG. 2

a k-dimensional pyramid may be expressed as a linear combination of the $k + 1$ vectors to the pyramid vertices. The expansion coefficients are made unique by requiring that they sum to one. Since A, Y, F, and B, X, F are collinear triples, we have

$$\vec{X} = X_B\vec{B} + X_F\vec{F} \tag{2}$$

$$\vec{Y} = Y_A\vec{A} + Y_F\vec{F} \tag{3}$$

$$X_B + X_F = Y_A + Y_F = 1 \tag{4}$$

And since F lies in R, we have

$$\vec{F} = F_C\vec{C} + F_D\vec{D} + \cdots + F_E\vec{E} \tag{5}$$

$$F_C + F_D + \cdots + F_E = 1. \tag{6}$$

Now using (2), (3), and (5) we see

$$\frac{X_C}{X_D} = \frac{X_F F_C}{X_F F_D} = \frac{F_C}{F_D} = \frac{Y_C}{Y_D} \tag{7}$$

Equation (7) says that in the resolution of base points into linear combinations of cell vertices, where 2 vertices enter the calculation of 2 base points, their "relative weights" must be equal. Choose masses M_C, M_D, M_E so that F is the *cm* of cell $C, D, \ldots E$. Then choose M_B so that X is the *cm* of cell $B, C, D, \ldots E$, and choose M_A so that Y is the *cm* of cell $B, C, D, \ldots E$. Now all the masses are assigned. Let Z be the base point opposite C in cell $A, B, D, \ldots E$. From (7) and the choice of Y as its cells *cm*, we have

$$\frac{Z_A}{Z_D} = \frac{Y_A}{Y_D} = \frac{M_A}{M_D}, \tag{8}$$

which generalizes to

$$Z_A : Z_B : Z_D \cdots : Z_E = M_A : M_B : M_D \cdots : M_E \tag{9}$$

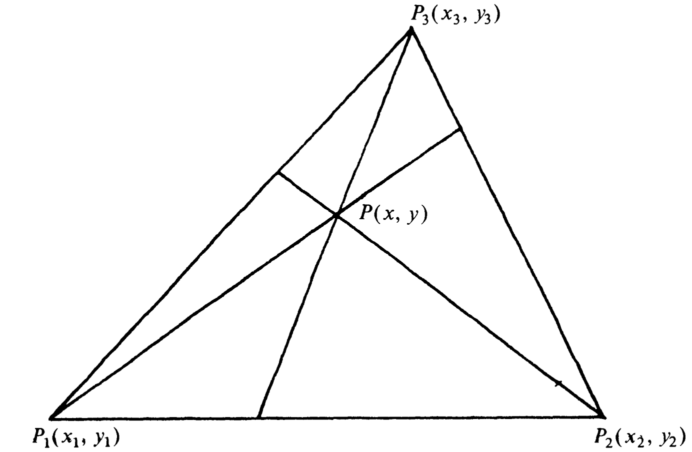


FIG. 3

Thus Z is also its cell's *cm*. Hence the sought for mass assignment exists and is necessary.

It is possible to determine the vertex masses for any Ceva configuration. Consider first the plane triangle (see FIGURE 3). Let (x, y) be the Ceva (cm) point. Suppose the total mass is 1. Then

$$x = M_1x_1 + M_2x_2 + M_3x_3 \tag{10}$$

$$y = M_1y_1 + M_2y_2 + M_3y_3 \tag{11}$$

$$1 = M_1 + M_2 + M_3. \tag{12}$$

We may solve this system by determinants to get

$$M_1 : M_2 : M_3 = \tag{13}$$

$$\begin{vmatrix} x & x_2 & x_3 \\ y & y_2 & y_3 \\ 1 & 1 & 1 \end{vmatrix} : \begin{vmatrix} x_1 & x & x_3 \\ y_1 & y & y_3 \\ 1 & 1 & 1 \end{vmatrix} : \begin{vmatrix} x_1 & x_2 & x \\ y_1 & y_2 & y \\ 1 & 1 & 1 \end{vmatrix} = \Delta PP_2P_3 : \Delta PP_1P_3 : \Delta PP_1P_2$$

This result may be generalized by stating that in any N-dimensional Ceva pyramid the mass assignment is such that the mass at any vertex is proportional to the "volume" of the opposite N-dimensional pyramid determined by the remaining N vertices and the Ceva point. In TABLE 1 and TABLE 2 are shown the proper mass

TABLE 1

Triangle Cevian	$M_a : M_B : M_C$
medians	$1:1:1$
altitudes	$\tan A : \tan B : \tan C$
angle bisectors	$\sin A : \sin B : \sin C$ $= a : b : c$ (sides)
to incircle tangency pts	$\tan A/2 : \tan B/2 : \tan C/2$
thru circumcenter ($\perp$ bisectors meet here)	$\sin 2A : \sin 2B : \sin 2C$ $= \tan B + \tan C : \tan A + \tan C : \tan A + \tan B$

TABLE 2

Pyramid Cevians	$M_A : M_B : M_C : M_D$
medians	$1:1:1:1$
angle bisectors (meet at center of insphere)	$a:b:c:d$ (face areas)

assignments for some of the classic Ceva configurations. These may be used to prove theorems in a straightforward manner. For example, by observing that collinear Ceva points are produced by linearly related mass assignments, the collinearity of the centroid, circumcenter, and orthocenter of any triangle on the Euler line is immediate. Another demonstration made easy using the *cm*-concurrence viewpoint shows the necessary and sufficient condition for the concurrence of all the altitudes of an N-dimensional pyramid. That condition is that nonadjacent edges be orthogonal.

REFERENCE

1. H. S. M. Coxeter and S. L. Greitzer, Geometry Revisited, Random House, 1967.

A Sequence of Coin Toss Variables for which the Strong Law Fails

S. RAMAKRISHNAN
Department of Mathematics, University of Miami, Coral Gables, FL 33124

W. D. SUDDERTH*
Department of Theoretical Statistics, University of Minnesota, Minneapolis, MN 55455

Introduction. Let $X = \{X_1, X_2, \dots\}$ be a sequence of $\{0,1\}$-valued random variables defined on a probability space $(\Omega, \mathscr{F}, P)$. Call X a *coin toss sequence* if

$$P[X_1 = i_1, X_2 = i_2, \dots, X_k = i_k] = 2^{-k} \tag{1}$$

for every finite sequence $i_1, \dots, i_k$ of 0's and 1's. It follows easily from (1) that $X_1, X_2, \dots$ are independent and identically distributed.

Define $S_n = X_1 + \cdots + X_n$ for $n = 1, 2, \dots$. The weak and the strong laws of large numbers for the sequence X are, respectively,

$$\lim_{n\to\infty} P[|S_n/n - 1/2| \leqq \varepsilon] = 1 \quad \text{for every } \varepsilon > 0. \tag{2}$$

and

$$P\left[\lim_{n\to\infty} S_n/n = 1/2\right] = 1. \tag{3}$$

*Research supported by National Science Foundation grant DMS-8421208.

Both (2) and (3) are classical results if $(\Omega, \mathscr{F}, P)$ is a conventional *countably additive probability space* in which $\mathscr{F}$ is a σ-field of subsets of Ω and P is a countably additive probability measure on $\mathscr{F}$.

Suppose instead that $(\Omega, \mathscr{F}, P)$ is a *finitely additive probability space* in which $\mathscr{F}$ is a field of subsets of Ω and P is a finitely additive probability measure on $\mathscr{F}$. The weak law remains true. Indeed the usual proof that (1) implies (2) does not rely on countable additivity. However, it is well known to students of finite additivity that the strong law need not hold.

At least two recent papers have given examples in which the strong law fails (Dubins and Freedman [**2**], Kumar and Fine [**5**]). The examples presented in these papers are nonconstructive and rely on more axioms than those of ZF (ZF is the Zermelo-Fraenkel set theory without the axiom of choice). The example presented in the next section is completely natural and constructive. It has the property that the convergence set $C \equiv [\lim_{n\to\infty} S_n/n = 1/2]$ is empty so that no extension of P is necessary to see that (3) fails. A slight refinement gives a constructive example for which the set $D \equiv [S_n/n \text{ converges}]$ is empty. Thus finitely additive spaces can be used to model bounded, stationary sequences for which averages fail to converge. Real world examples of such phenomena have been reported by Kumar and Fine [**5**] who suggest nonadditive models.

The Example. Let $\Omega = \{1, 2, \dots\}$ be the set of positive integers and let $\mathscr{F}$ be the collection of all finite unions of congruence classes of the form

$$A(n, r) = \{\omega \in \Omega | \omega \quad \text{has remainder } r \text{ when divided by } n\},$$

where $n = 1, 2, \dots$; $r = 0, 1, \dots, n-1$. It is easy to check that $\mathscr{F}$ is an algebra.

It is also straightforward to verify that there is a unique finitely additive probability measure P on $\mathscr{F}$ satisfying

$$P(A(n, r)) = 1/n \quad \text{for every set } A(n, r).$$

The space $(\Omega, \mathscr{F}, P)$ is familiar to number theorists who consider certain extensions of P. It was also mentioned by Dubins and Savage [**3**] in their seminal book on finite additivity and gambling.

To define the coin toss sequence $X = \{X_1, X_2, \dots\}$, write $\omega \in \Omega$ in its unique binary expansion

$$\omega = a_1 \times 2^0 + a_2 \times 2^1 + \cdots$$

and set

$$X_k(\omega) = a_k.$$

For a sequence $i_1, \dots, i_k$ of 0's and 1's let

$$i_k \dots i_1 = i_1 \times 2^0 + \cdots + i_k \times 2^{k-1}.$$

It can then be verified that

$$\{\omega | X_1(\omega) = i_1, \dots, X_k(\omega) = i_k\} = A(2^k, i_k \dots i_1). \tag{5}$$

(For example, $[X_1 = 0]$ is the set of even numbers.) Property (1) now follows from (4) and (5). Thus X is a coin toss sequence.

Now, for every $\omega \in \Omega$, there is a positive integer k_0 such that $a_k = X_k(\omega) = 0$ for $k \geqq k_0$. Hence, $S_n(\omega)/n$ converges to 0 for every ω.

Next we will construct a coin toss sequence $Y = \{Y_1, Y_2, \dots\}$ such that $(Y_1(\omega) + \cdots + Y_n(\omega))/n$ fails to converge, for every ω. To define Y, let $j_1, j_2, \dots$ be an infinite sequence of 0's and 1's such that $(j_1 + \cdots + j_n)/n$ does not converge. Let

$$\begin{aligned} Y_n &= X_n && \text{if } j_n = 0, \\ &= 1 - X_n && \text{if } j_n = 1. \end{aligned}$$

It is easily checked that Y is a coin toss sequence. Also, for every ω, $Y_n(\omega)$ is eventually equal to j_n because $X_n(\omega)$ is eventually equal to 0.

By the way, an example quite similar to the above can be constructed by taking Ω to be the rational numbers in $[0,1]$, $\mathscr{F}$ to be all finite unions of intervals of rationals and P to be the unique probability on $\mathscr{F}$ such that $P(I) =$ "length" of I for every interval I of rationals. The random variables $X_1, X_2, \dots$ correspond to a binary expansion as before.

Remarks. We have shown this example to a number of our friends who are conventional, countably additive probabilists. Most of them are amused by the example and consider it to be evidence of the perversities of finite additivity. However, we view it as evidence of the arbitrariness of the assumption of countable additivity since we know of no simpler or more elementary model of a coin toss sequence. It seems to us a mistake to ban such an example from consideration. (A recent editor of the *Annals of Probability* ruled that "finitely additive probability is not probability.")

There is a natural class of finitely additive probability measures which includes the countably additive measures and for which the classical strong laws of probability theory do hold. (See Purves and Sudderth **[6]**, Chen **[1]**, Ramakrishnan **[7]**, and Karandikar **[4]**.)

REFERENCES

1. R. Chen, On almost sure convergence in a finitely additive setting, *Z. Wahrsch. Verw. Gebiete*, 37 (1977) 341–356.
2. L. E. Dubins, and D. A. Freedman, Exchangeable processes need not be mixtures of independent, identically distributed random variables, *Z. Wahrsch. Verw. Gebiete*, 48 (1979) 115–132.
3. L. E. Dubins and L. J. Savage, Inequalities for Stochastic Processes, Dover, 1976.
4. R. L. Karandikar, A general principle for limit theorems in finitely additive probability, *Trans. Amer. Math. Soc.*, 273 (1982) 541–550.
5. A. Kumar and T. L. Fine, Stationary lower probabilities and unstable averages, *Z. Wahrsch. Verw. Gebiete*, 69 (1985) 1–17.
6. R. A. Purves and W. D. Sudderth, Some finitely additive probability, *Ann. Probab.*, 4 (1976) 259–276.
7. S. Ramakrishnan, Finitely additive Markov chains, *Trans. Amer. Math. Soc.*, 265 (1981) 247–272.

THE TEACHING OF MATHEMATICS

EDITED BY MELVIN HENRIKSEN AND STAN WAGON

A Simple Test for the *n*th term of a Series to Approach Zero

JONATHAN LEWIN
Department of Mathematics, Kennesaw College, Marietta, GA 30061

MYRTLE LEWIN
Department of Mathematics, Agnes Scott College, Decatur, GA 30030

Using Stirling's formula, one may see at once that if $a_n = (2n)!/4^n(n!)^2$, then a_n is of the order of $1/\sqrt{n}$, and one may conclude from the alternating series test that the series $\Sigma(-1)^n a_n$ is conditionally convergent. At an elementary level, however, the convergence of the latter series may be a little more difficult to obtain. Since $a_{n+1}/a_n = (2n+1)/(2n+2) < 1$ for each n, it is clear that the sequence (a_n) is decreasing, but it is not immediately obvious within the environment of a typical calculus course that $a_n \to 0$ as $n \to \infty$. For this purpose, one might use the following simple result which takes a leaf out of the theory of infinite products:

THEOREM. *Suppose (a_n) is a decreasing sequence of positive numbers and for each natural number n, define $b_n = 1 - a_{n+1}/a_n$. Then the sequence (a_n) converges to zero if and only if the series Σb_n diverges.*

Proof. We note first that unless $b_n \to 0$ as $n \to \infty$, both of the series Σb_n and $\Sigma \log(1 - b_n)$ diverge. On the other hand, if $b_n \to 0$, then $b_n/(-\log(1 - b_n)) \to 1$ as $n \to \infty$, and it follows from the limit comparison test that Σb_n diverges if and only if $\Sigma \log(1 - b_n)$ diverges. We note also that since $0 < b_n < 1$, we have $\log(1 - b_n) < 0$ for each n.

Now since $1 - b_n = a_{n+1}/a_n$ for every n, it is clear that $a_n = a_1(1 - b_1)(1 - b_2)(1 - b_3) \cdots (1 - b_{n-1})$ for each $n \geqslant 2$, and we therefore conclude that $a_n \to 0$ iff $\log a_n \to -\infty$ iff $\log a_1 + \Sigma_{i=1}^{n} \log(1 - b_i) \to -\infty$ iff $\Sigma \log(1 - b_n)$ diverges iff Σb_n diverges.

Returning now to the above example, we see that $b_n = 1/(2n+2)$ for each n, and the obvious divergence of Σb_n implies that $a_n \to 0$. The same technique gives an easy proof of the convergence of such series as $\Sigma((-1)^n n^n/e^n n!)$, and the series $\Sigma\binom{\alpha}{n}$ of binomial coefficients with $\alpha > -1$.

Universal Topological Spaces

K. D. MAGILL, JR.
Department of Mathematics, SUNY, Buffalo, NY 14214

Let $U = \{a, b, c\}$ and let $\mathscr{T}_1 = \{U, \phi, \{a\}\}$. It has been known for a long time that U with the topology $\mathscr{T}_1$ is a *universal topological space* in the sense that any topological space whatsoever is homeomorphic to a subspace of some topological

power U^α of U. Indeed, this is an easy consequence of the well-known Embedding Lemma of [**1**, p. 116] (which appears in [**4**] as Cor. 8.15). The main problem we consider here is that of finding all universal topological spaces. We present two characterizations whose proofs involve only elementary notions and techniques. Thus this problem is appropriate for the students in a beginning topology course after they have been presented with the Embedding Lemma. We now state a version of the Lemma which will be convenient for our purposes here.

EMBEDDING LEMMA. *Let X be any topological space whatsoever and let $\{Y_\alpha\}_{\alpha\in\Lambda}$ be any collection of topological spaces. For each $\alpha \in \Lambda$, let f_α be a continuous function from X to Y_α and suppose the collection of functions $\{f_\alpha\}_{\alpha\in\Lambda}$ separates points and also separates points and closed sets. Then the evaluation map e from X into $\Pi\{Y_\alpha\}_{\alpha\in\Lambda}$ defined by $e(x)_\alpha = f_\alpha(x)$ for each $\alpha \in \Lambda$ is an embedding.*

Traditional applications of the Embedding Lemma include characterizing completely regular Hausdorff spaces as subspaces of compact Hausdorff spaces and, indeed, producing the Stone-Čech compactification of a completely regular Hausdorff space as well as deducing the metrization theorems. As we mentioned previously, another very simple application of the Embedding Lemma is to prove that U is a universal space. One need only verify that for any space X, the family $C(X, U)$ of all continuous maps from X into U separates points and also separates points and closed sets. For distinct points $x, y \in X$, define $f(x) = b$ and $f(z) = c$ for $z \neq x$. For a closed subset $F \subseteq X$ and $w \notin F$, define $g(z) = b$ for $z \in F$ and $g(z) = a$ for $z \notin F$. Both f and g are continuous and $g(w) \notin \operatorname{cl} g[F]$. Thus, maps of the type f separate points while those of the type g separate points and closed sets.

It is not too much to ask students in a beginning topology course to produce such maps. Finding a universal space is more difficult but my experience has been that the better students can also do that although they often produce considerably more complicated spaces than is necessary. Some students may eventually need hints. Nevertheless, the exercise is an informative one in that it underscores, at an early stage, just how complicated things can get by taking products and subspaces.

To my knowledge, the statement that U is a universal space first appeared in print in [**3**] but it doesn't seem to be as well known as one as one might expect for such a surprising result. Incidentally, a nice historical discussion of various embedding results may be found in [**3**].

THEOREM 1. *X is a universal space if and only if it is not T_0 and it contains two points x and y such that some open set contains x but not y while every open subset that contains y also contains x.*

Proof. Suppose first that X is a universal space. Then X cannot be T_0 since products of T_0 spaces are T_0 and subspaces of T_0 spaces are also T_0. Now let $W = \{a, b\}$ and let W have the subspace topology induced by U. Then some topological power X^α of X must contain a copy Y of W. That is, $Y = \{p, q\}$ where the topology induced on Y is $\{\phi, Y, \{p\}\}$. Thus there exists an open subset G of X^α containing p but not q and, furthermore, every open subset of X^α that contains q also contains p. Then

$$p \in \pi_{\alpha_1}^{-1}[G_{\alpha_1}] \cap \pi_{\alpha_2}^{-1}[G_{\alpha_2}] \cap \cdots \cap \pi_{\alpha_N}^{-1}[G_{\alpha_N}] \subseteq G$$

and

$$q \notin \pi_{\alpha_1}^{-1}[G_{\alpha_1}] \cap \pi_{\alpha_2}^{-1}[G_{\alpha_2}] \cap \cdots \cap \pi_{\alpha_N}^{-1}[G_{\alpha_N}],$$

where for each α_i, π_{α_i} is the corresponding projection map and G_{α_i} is an open subset of $X_{\alpha_i} = X$. Evidently, for some α_i, $q_{\alpha_i} \notin G_{\alpha_i}$ (where q_{α_i} is the α_ith coordinate of q) while $p_{\alpha_i} \in G_{\alpha_i}$. Moreover, every open subset of X that contains q_{α_i} must also contain p_{α_i} since every open subset of X^α containing q must also contain p.

For the converse, suppose that X is not T_0 and that it has two points x and y as described in the statement of the theorem. Since X is not T_0, it also contains two points a and b such that $a \neq b$ but both a and b are contained in precisely the same open subsets of X. Next, let $Z = \{(a, y), (b, y), (a, x)\}$ and let Z have the subspace topology induced by $X \times X$. Let G be an open subset of X containing x but not y. Then $Z \cap (X \times G) = \{(a, x)\}$ is open in Z. Now suppose V is any open subset of $X \times X$ and $(a, y) \in Z \cap V$. Then

$$(a, y) \in Z \cap (G_1 \times G_2) \subseteq Z \cap V,$$

where G_1 and G_2 are open subsets of X. Since a and b belong to precisely the same open subsets of X, $(b, y) \in Z \cap V$. Moreover, $y \in G_2$ implies $x \in G_2$ so that $(a, x) \in Z \cap V$. Similarly, $(b, y) \in Z \cap V$ implies that both (a, y) and (a, x) belong to $Z \cap V$ as well. Evidently the topology induced by Z is $\{\varnothing, Z, \{(a, x)\}\}$. That is, $X \times X$ contains a copy of U which is a universal space as we verified previously. It is now immediate that X itself must be a universal space.

The next two corollaries follow easily from Theorem 1 and we omit their proofs.

COROLLARY 2. *If X is a universal space then X must have at least three points.*

COROLLARY 3. *Up to homeomorphism there are exactly two universal topological spaces with exactly three points and each has precisely one nonempty, proper, open subset. Consequently, in the one case it is a singleton and in the other, a pair of points.*

We proved the second half of Theorem 1 by showing that if the space X satisfies certain conditions then $X \times X$ contains a copy of $(U, \mathscr{T}_1)$ where $\mathscr{T}_1 = \{\varnothing, U, \{a\}\}$. We have also observed in Corollary 3 that $(U, \mathscr{T}_2)$ is a universal space where $\mathscr{T}_2 = \{\varnothing, U, \{a, b\}\}$. One might be tempted to try to show that if X satisfies those conditions then X itself must contain a copy of either $(U, \mathscr{T}_1)$ or $(U, \mathscr{T}_2)$ but as the following example shows, this is not possible.

Example. Let $V = \{a, b, c, d\}$ and

$$\mathscr{T}_3 = \{\varnothing, V, \{a\}, \{a, b\}, \{c, d\}, \{a, c, d\}\}.$$

It follows immediately from Theorem 1 that $(V, \mathscr{T}_3)$ is a universal space and one can use that theorem to check that no proper subspace of $(V, \mathscr{T}_3)$ is a universal space.

In a sense, then, the three spaces $(U, \mathscr{T}_1)$, $(U, \mathscr{T}_2)$, and $(V, \mathscr{T}_3)$ are all minimal universal spaces. Of course, any space that contains a copy of any one of these spaces is also a universal space. Our next theorem shows that the converse is true as well and this results in another quite different characterization of universal spaces.

For a subspace Y of a topological space X, the symbol $\mathscr{T}_Y$ will denote the topology induced on Y.

THEOREM 4. *A topological space X is a universal space if and only if it contains a copy of one of the spaces $(U, \mathscr{T}_1)$, $(U, \mathscr{T}_2)$, or $(V, \mathscr{T}_3)$.*

Proof. We have already discussed the fact that the condition is sufficient so we prove only necessity. By Theorem 1, X contains two distinct points a and b that belong to precisely the same open sets and also two points c and d such that some open set contains c but not d while every open subset containing d also contains c.

Case 1.

$$c = a \quad \text{or} \quad c = b.$$

We may assume that $c = a$. Then $d \neq a, b$. Let G be the open set containing a but not d and let $Y = \{a, b, d\}$. Then $Y \cap G = \{a, b\}$ is open in Y and moreover is the only proper open subset since every open set containing d also contains a and hence b as well. Thus $\mathscr{T}_Y = \{Y, \varnothing, \{a, b\}\}$ and $(Y, \mathscr{T}_Y)$ is homeomorphic to $(U, \mathscr{T}_2)$.

Case 2.

$$d = a \quad \text{or} \quad d = b.$$

We may assume $d = a$. Then $c \neq a, b$. Let $Y = \{a, b, c\}$ and let G be an open subset of X that contains c but not $d = a$. Then $b \notin G$ and $G \cap Y = \{c\}$. Every open subset of X that contains either a or b must contain the other as well and hence also c since $d = a$. In this case $\mathscr{T}_Y = \{Y, \varnothing, \{c\}\}$ and $(Y, \mathscr{T}_Y)$ is homeomorphic to $(U, \mathscr{T}_1)$.

Case 3.

$$c \neq a, b \quad \text{and} \quad d \neq a, b.$$

In this case a, b, c and d are all distinct and we let $Y = \{a, b, c, d\}$ with induced topology $\mathscr{T}_Y$. It follows that a and b are contained in precisely the same open subsets of Y and there exists an open subset of Y containing c but not d while every open subset containing d contains c as well. Now let E be the smallest open subset of Y containing a and b, let G be the smallest open subset of Y containing c, and let H be the smallest open subset of Y containing d. There are three possibilities for E, two for G, and two for H. They are as follows:

$(E1)$ $E = Y$	$(G1)$ $G = \{a, b, c\}$	$(H1)$ $H = Y$
$(E2)$ $E = \{a, b, c\}$	$(G2)$ $G = \{c\}$	$(H2)$ $H = \{c, d\}$
$(E3)$ $E = \{a, b\}$		

Now $(E1)$ together with $(G1)$ is contradictory as is $(G1)$ and $(H2)$ together. In the latter instance, for example, if $\{a, b, c\}, \{c, d\} \in \mathscr{T}_Y$ then $\{c\} \in \mathscr{T}_Y$ which contradicts the fact that $\{a, b, c\}$ is the smallest open set containing c. We now list all the other possibilities and simply observe that for each of these, $(Y, \mathscr{T}_Y)$ and hence the space X contains a subspace homeomorphic to either $(U, \mathscr{T}_1)$, $(U, \mathscr{T}_2)$, or $(V, \mathscr{T}_3)$. We express the fact that $(Y, \mathscr{T}_Y)$ is homeomorphic to $(Z, \mathscr{T}_Z)$ by simply writing $(Y, \mathscr{T}_Y) \approx (Z, \mathscr{T}_Z)$.

$(E1)\,\&\,(G2)\,\&\,(H1)$	Let $Z = Y - \{d\}$.	Then $(Z, \mathscr{T}_Z) \approx (U, \mathscr{T}_1)$.
$(E1)\,\&\,(G2)\,\&\,(H2)$	Let $Z = Y - \{c\}$.	Then $(Z, \mathscr{T}_Z) \approx (U, \mathscr{T}_1)$.

$(E2) \& (G1) \& (H1)$	Let $Z = Y - \{a)$.	Then $(Z, \mathscr{T}_Z) \approx (U, \mathscr{T}_2)$.
$(E2) \& (G2) \& (H1)$	Let $Z = Y - \{d\}$.	Then $(Z, \mathscr{T}_Z) \approx (U, \mathscr{T}_1)$.
$(E2) \& (G2) \& (H2)$	Let $Z = Y - \{d\}$.	Then $(Z, \mathscr{T}_Z) \approx (U, \mathscr{T}_1)$.
$(E3) \& (G1) \& (H1)$	Let $Z = Y - \{d\}$.	Then $(Z, \mathscr{T}_Z) \approx (U, \mathscr{T}_2)$.
$(E3) \& (G2) \& (H1)$	Let $Z = Y - \{c\}$.	Then $(Z, \mathscr{T}_Z) \approx (U, \mathscr{T}_2)$.
$(E3) \& (G2) \& (H2)$	Then $(Y, \mathscr{T}_Y) \approx (V, \mathscr{T}_3)$.	

This concludes the proof of the theorem.

Some Further Exercises

One can have students investigate related problems by defining a space X to be **C**-universal (where **C** is a prescribed class of topological spaces) if $X \in \mathbf{C}$ and each $Y \in \mathbf{C}$ is homeomorphic to a subspace of some topological power X^α of X. Let **Tz** denote the class of all T_0 spaces, **Ty** the class of all Tychonoff spaces, and **Zd** the class of all 0-dimensional Hausdorff spaces. A 0-dimensional space here is one that has a basis of sets that are both open and closed. Exercises 1, 2 and 3 are straightforward applications of the Embedding Lemma and the techniques used in this paper.

Exercise 1. Prove that the following statements about a topological space X are equivalent.

(1.1) X is **Tz**-universal.

(1.2) X is T_0 and not T_1.

(1.3) X is T_0 and contains a copy of the space $(Y, \mathscr{T})$ where $Y = \{a, b\}$ and $\mathscr{T} = \{Y, \varnothing, \{a\}\}$.

Exercise 2. Prove that a topological space is **Ty**-universal if and only if it is a Tychonoff space and contains an arc.

Exercise 3. Prove that a topological space is **Zd**-universal if and only if it is a 0-dimensional Hausdorff space and contains more than one point.

Of course, Exercises 2 or 3 may be inappropriate at this particular point, depending upon what topics have been covered previously. One might then assign the following problem.

Exercise 4. Is there a result about T_1 spaces similar in nature to the result of Exercise 1?

If, after a time, students show no success in dealing with Exercise 4 it can be suggested that they check **[2]**.

REFERENCES

1. J. L. Kelley, General Topology, Van Nostrand, New York, 1955.
2. S. Mrowka, On universal spaces, *Bull Acad. Polon. Sci.*, cl. III, 5 (1957) 479–481.
3. ______, Further results on E-compact spaces, I, *Acta Math.*, 120 (1968) 161–185.
4. S. Willard, General Topology, Addison-Wesley, Reading, MA., 1968.

3. W. P. Galvin, Matrices with "custom-built" eigenspaces, this MONTHLY, 91(1984) 308–309.
4. J. M. Ortega, Letter to the editor, this MONTHLY, 92(1985) 526.
5. L. J. Paige and J. D. Swift, Elements of Linear Algebra, Blaisdell, N.Y., 1961.
5. J.-C. Renaud, Matrices with integer entries and integer eigenvalues, this MONTHLY, 90(1983) 202–203.

A New Algorithm for Computing the Rank of a Matrix

LARRY J. GERSTEIN
Department of Mathematics, University of California, Santa Barbara, CA 93106

To determine the rank of a matrix A and a basis for its row space, students are usually instructed to reduce A to echelon form by using elementary row operations. Then the number of nonzero rows of the echelon matrix is the rank of A, and the rows of the echelon matrix constitute a basis for the row space of A.

While this is quite correct, in practice the computations can become unwieldy. For example, if A is an integer matrix there are two standard ways to proceed. In the first approach we start by using elementary row operations (based on the division algorithm) to eliminate all but one entry in the first column of A. In the second approach we first divide the top row by the left-hand entry a_{11} (assuming $a_{11} \neq 0$), then subtract suitable multiples of the new top row from all the others; in either case the computation reduces matters to considering an $(m-1) \times (n-1)$ matrix, whereupon the process is repeated.

The first approach has the advantage that all computations are performed with integers, but a great many row operations are likely to be needed, and there may be some uncertainty as to exactly which row operations should be used in order to minimize computation. In the second approach denominators will immediately appear on the scene (unless A has been specially designed to make computations easy, say by having lots of rows with leading entry 1 involved in the calculations), and the potential for error then increases dramatically. Indeed, many texts make the appearance of nontrivial denominators especially likely by including the requirement that each nonzero row of an "echelon" matrix have 1 as its first nonzero entry. That requirement is irrelevant for rank and row space determinations. (Of course, a leading 1 *is* important if the reduction is being done in order to solve a system of linear equations.)

Our purpose here is to present a rank algorithm in which division is never needed. Like the usual procedures, this one yields an echelon matrix (not necessarily with a leading 1 in its nonzero rows) that is row-equivalent to the original, and whose nonzero rows constitute a basis for the row space of A. But here there will be no uncertainty about how to proceed, and denominators will not arise in the computations.

First we fix some notation. Let F be a field, and let $M_{m\times n}(F)$ denote the set of $m \times n$ matrices over F. For any matrix $A = (a_{ij}) \in M_{m\times n}(F)$ and any pair of indices i, j satisfying $2 \leqslant i \leqslant m$ and $2 \leqslant j \leqslant n$, we define the 2×2 subdeterminant

$$d_{ij} = \begin{vmatrix} a_{11} & a_{1j} \\ a_{i1} & a_{ij} \end{vmatrix} = a_{11}a_{ij} - a_{i1}a_{1j}.$$

The following result provides the key to our procedure.

THEOREM. *Let $A = (a_{ij}) \in M_{m\times n}(F)$, and suppose $a_{11} \neq 0$. Then*

$$\operatorname{rank} A = 1 + \operatorname{rank}\begin{pmatrix} d_{22} & \cdots & d_{2n} \\ \vdots & & \\ d_{m2} & \cdots & d_{mn} \end{pmatrix}.$$

Proof. Write $\sim$ for row equivalence of matrices. We first multiply rows 2 through m by a_{11}, obtaining

$$A \sim \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{11}a_{21} & a_{11}a_{22} & \cdots & a_{11}a_{2n} \\ \vdots & & & \\ a_{11}a_{m1} & a_{11}a_{m2} & \cdots & a_{11}a_{mn} \end{pmatrix}.$$

Now we take this new matrix and, for each row index i satisfying $2 \leqslant i \leqslant m$, we subtract a_{i1} times row 1 from row i. This yields the row equivalence

$$A \sim \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & d_{22} & \cdots & d_{2n} \\ \vdots & & & \\ 0 & d_{m2} & \cdots & d_{mn} \end{pmatrix}.$$

Because $a_{11} \neq 0$, the first row of this last matrix is not a linear combination of the other rows. The conclusion follows. □

An algorithm for rank computation. First note that if a given nonzero matrix $A = (a_{ij})$ has a leading column of zeros, then the rank is unchanged when that column is deleted; thus we may assume the first column to be nonzero. Moreover, by using an elementary row operation (if necessary) we can assume without loss of generality that $a_{11} \neq 0$. Making use of these observations as needed, we now apply the theorem recursively until we obtain a matrix with only one row or column, and then the solution is evident. (In practice, the rank of a matrix with only two rows or two columns is usually evident.)

Examples.

(i) $$\operatorname{rank}\begin{pmatrix} 3 & -8 & 7 \\ 5 & -4 & 9 \\ 2 & 3 & 6 \end{pmatrix} = 1 + \operatorname{rank}\begin{pmatrix} 28 & -8 \\ 25 & 4 \end{pmatrix} = 2 + \operatorname{rank}(312) = 3.$$

(ii) $$\operatorname{rank}\begin{pmatrix} 4 & 3 & -5 & 6 \\ 6 & 2 & 0 & 2 \\ 3 & 5 & -12 & 5 \\ 2 & 2 & -4 & 2 \end{pmatrix} = 1 + \operatorname{rank}\begin{pmatrix} -10 & 30 & -28 \\ 11 & -33 & 2 \\ 2 & -6 & -4 \end{pmatrix}$$

$$= 2 + \operatorname{rank}\begin{pmatrix} 0 & 288 \\ 0 & 96 \end{pmatrix} = 3.$$

Remarks. (a) The given matrix A is row-equivalent to the echelon matrix constructed by stacking up the top rows of the matrices appearing in the rank computations, with the understanding that leading zeros need to be prefixed to the

rows that have fewer than n entries. For example, the matrix A in Example (i) is row-equivalent to the echelon matrix

$$\begin{pmatrix} 3 & -8 & 7 \\ 0 & 28 & -8 \\ 0 & 0 & 312 \end{pmatrix},$$

and the rows of this matrix constitute a basis of the row space of A. (Of course in this example A is nonsingular, so its *own* rows are also a basis for the row space.) Similarly, in Example (ii) the nonzero rows of the echelon matrix

$$\begin{pmatrix} 4 & 3 & -5 & 6 \\ 0 & -10 & 30 & -28 \\ 0 & 0 & 0 & 288 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

constitute a basis for the row space of A. In building up the echelon matrices in this way, one must be careful to reinsert any columns of zeros that were deleted during the rank computations.

(b) Let A be an $n \times n$ matrix. Each reduction from a $k \times k$ matrix to a $(k-1) \times (k-1)$ matrix requires $2(k-1)^2$ multiplications. Adding these numbers for $k = 2, 3, \ldots, n$, gives a total of $(2n^3 - 3n^2 + n)/3 \approx 2n^3/3$ multiplications, which has the same order of magnitude as the number required to compute the determinant of A (see [**1**, pp. 479–480]).

REFERENCES

1. Donald E. Knuth, The Art of Computer Programming, Volume 2 (second edition), Addison-Wesley, Reading, Mass., 1981.

PROBLEMS AND SOLUTIONS

EDITED BY PAUL T. BATEMAN, HAROLD G. DIAMOND, KENNETH B. STOLARSKY (ADVANCED PROBLEMS), AND DOUGLAS B. WEST (ELEMENTARY PROBLEMS)

For instructions about submitting ***proposed*** *problems for publication in this department see the inside front cover. Please include soltuions, revelant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

Solutions *should be sent to the address given on the inside front cover. Two copies suffice.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so múch the better. If you discover that a Monthly problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.

For instructions about submitting solutions of Problems, which should be mailed before April 30, 1989, see the inside front cover. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label if an acknowledgment is desired.

ELEMENTARY PROBLEMS

E 3295. *Proposed by A. W. Goodman, University of South Florida, Tampa, FL.*

A function f on the open unit disc $E = \{z : |z| < 1\}$ is said to be *univalent* if it is analytic on E and takes no value more than once. A function f on E is said to be *infinite-valent* if it is analytic on E and takes some value infinitely often.

(a) Is there a univalent function f on E such that f' is nonconstant and infinite-valent on E?

(b) Is there an infinite-valent function g on E such that g' is univalent on E?

E 3296. *Proposed by Robert M. Young, Oberlin College, Oberlin, OH.*

Given four arbitrary points in the plane, describe a procedure for constructing the thinnest annulus containing all four points. (An infinite strip is regarded as an extreme case of an annulus.) The thickness of an annulus is the difference between the two radii (or the width of the infinite strip).

E 3297. *Proposed by Gunnar Blom, University of Lund and Lund Institute of Technology, Sweden.*

If m and k are positive integers and $m + 2k = pr$, where $p \geqslant k$ is an odd prime and r is a positive integer, prove that

$$2^{m+1}\sum_{j=0}^{k-1}(-1)^j\binom{m+j}{j} \equiv 2^{r-1} (\mathrm{mod}\ p).$$

E 3298. *Proposed by Marvin Marcus and Claire Pesce, University of California, Santa Barbara, CA.*

If A and B are square matrices, we define the Kronecker product $A \otimes B$ as the matrix

$$\begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1}B & a_{n2}B & \cdots & a_{nn}B \end{bmatrix},$$

where

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}.$$

Let $A_1, A_2, \ldots, A_m$ be square matrices over a field and let $K = A_1 \otimes A_2 \otimes \cdots \otimes A_m$. Prove that K is lower triangular if and only if the following two conditions are satisfied: (a) A_1 is lower triangular; (b) if there exists $p < m$ such that A_{p+1} is not lower triangular but all of $A_1, A_2, \ldots, A_p$ are lower triangular, then at least one of the matrices $A_1, A_2, \ldots, A_p$ has zero main diagonal.

E 3299. *Proposed by Jihad Yamout, Ohio University, Athens, OH.*

Suppose ABCD is a plane quadrilateral with no two sides parallel. Put $E = \overleftrightarrow{AB} \cap \overleftrightarrow{CD}$ and $F = \overleftrightarrow{AD} \cap \overleftrightarrow{BC}$. If M, N, P are the midpoints of AC, BD, EF, respectively, and $\overrightarrow{AE} = a\overrightarrow{AB}$, $\overrightarrow{AF} = b\overrightarrow{AD}$, where a and b are nonzero real numbers, prove that $\overrightarrow{MP} = ab\overrightarrow{MN}$.

E 3300. *Proposed by H. Das Gupta, Kanchrapara, West Bengal, India.*

Let N be the set of natural numbers. Prove that there is exactly one mapping f from $N \times N$ to N having the following three properties for $x, y \in N$:

(i) $f(x, y) = f(y, x)$,
(ii) $f(x, x) = x$,
(iii) $(y - x)f(x, y) = yf(x, y - x)$ if $y > x$.

SOLUTIONS OF ELEMENTARY PROBLEMS

Palindromes with Two Letters

E 3156 [1986, 482]. *Proposed by Raphael M. Robinson, University of California, Berkeley.*

Suppose that r, s, t are integers with $r \geqslant 0$, $s \geqslant 0$, $t = r + s \geqslant 2$. Is there a word W of length t in the alphabet $\{a, b\}$ such that $W = AB = Cab$, where A, B, C are palindromes, and the lengths of A and B are r and s? Show that such a word W exists if and only if $r + 2$ is prime to $s - 2$, and that in this case it is unique.

Solution by Allan Pedersen, Søborg, Denmark. Given a word V, let $\overline{V}$ denote the reverse of V, and let $|V|$ denote the length of V. We need several statements about the solutions to word equations.

LEMMA 1. *If $|A| = |B| + 1$, the equation $A = Ba$ has a unique solution in which A and B are palindromes, namely $A = a^{|A|}$ and $B = a^{|B|}$. Also, if $|A| = |B| + 2$, the equation $A = Bab$ has a solution in which A and B are palindromes if and only if $|B|$ is*

odd, in which case the solution is unique and is $A = (ba)^{(|A|-1)/2}b$ *and* $B = (ba)^{(|B|-1)/2}b$.

Proof. We prove only the second claim, since the proof of the first is similar. The proof is by induction, with the claims holding by inspection for $|B| \leqslant 1$. Suppose $|B| \geqslant 2$. Since A and $\overline{B}$ are palindromes, we have $Bab = A = baB$. Hence $A = baB'ab$, where $B' = \overline{B}'$ and $B = B'ab$. Let $A' = B$. Then we must solve $A' = B'ab$, and the desired result follows from the induction hypothesis.

LEMMA 2. *The equation* $C = AabB$ *has a solution in which* A, B, C *are palindromes and* $|A| = r$, $|B| = s$ *if and only if* $\gcd(r+2, s+2) = 1$, *in which case the solution is unique.*

Proof. Since C is a palindrome, we may assume $r \geqslant s$. If $r = s$ there can be no solution, since $ab \neq ba$. If $r = s + 1$, then $A = \overline{A} = bB$; by Lemma 1, there is a unique solution. If $s = 0$, Lemma 1 shows there is a solution if and only if r is odd, in which case it is unique. For the remaining cases we have $r - s \geqslant 2$ and $s > 0$, and we proceed by induction on $r + s$, reducing to the cases above.

Since C is a palindrome, reversing C and deleting $|C| - |A| = s + 2$ letters from the end yields $A = BbaA'$, with $|A'| = r - s - 2 \geqslant 0$. By substitution, $C = BbaA'abB$, so $\overline{A'} = A'$. Thus $C = AabB$ has a solution if and only if $A = BbaA'$ has a solution in palindromes. In this equation we have $|A| = r$, $|B| = s$ and $|A'| = r - s - 2$. If $s < r - s - 2$ we can reverse the equation. By induction, the equation has a solution if and only if $\gcd(s+2, r-s) = 1$, which is true if and only if $\gcd(s+2, r+2) = 1$. Furthermore, if the shorter equation has a solution, it is unique. Since it yields a solution to the longer equation in a unique way, solutions to the longer equation are also unique. Thus Lemma 2 is proved.

Now we are ready to prove the claim of the problem. We proceed by induction on s. If $s = 0$ or $s = 1$, then the claim (including the primality condition) follows from Lemma 1. If $s = 2$, then $B = ab$, which is not a palindrome. Hence there is no solution, corresponding to the fact that $\gcd(r+2, 0) = r + 2$. If $s = 3$, then $B = bab$ since B is a palindrome. Now $Ab = C$, and by Lemma 1 there is always a unique solution, corresponding to the fact that $\gcd(r+2, 1) = 1$. Finally, suppose $s \geqslant 4$. Now $B = baB'ab$ where $B' = \overline{B}'$. Deleting two letters from the end of the equated words, we have $AbaB' = C$. By Lemma 2, we have a solution if and only if $\gcd(r+2, s-4+2) = \gcd(r+2, s-2) = 1$, in which case the solution is unique.

Editorial Comment. Most solutions were inductive. J. M. Cohen provided a solution using group theory, and R. B. Eggleton and P. J. Zwier provided solutions using graph theory.

Also solved by I. C. Bivens, J. M. Cohen, R. B. Eggleton (Australia), J. Fukuta (Japan), O. P. Lossers (The Netherlands), P. J. Zwier, and the proposer.

Some Jordan Curves and Convex Curves

E 3158 [1986, 482]. *Proposed by I. J. Schoenberg, Madison, WI.*

Consider the image C_n of the circle $z = e^{it} (0 \leqslant t \leqslant 2\pi)$ under the function $f(z) = z + az^n$, where a is a real constant. Show that:

(a) If $n \geqslant 2$, then C_n is a closed Jordan curve if and only if $|a| \leqslant 1/n$.

(b) If $n \geqslant 2$, then C_n is a closed convex curve if and only if $|a| \leqslant 1/n^2$.

Editorial Remark. The condition $n \geqslant 2$ was inadvertently omitted from (a) in the published statement of the problem.

Solution by Allan Pedersen, Søborg, Denmark. (a) If C_n has multiple points, i.e., if there exists a pair of distinct numbers z and w of modulus 1 such that $z + az^n = w + aw^n$, then

$$-1 = aw^{n-1}\{1 + z/w + (z/w)^2 + \cdots + (z/w)^{n-1}\},$$
$$|z| = |w| = 1, \qquad z \neq w \tag{*}$$

and, hence, $1 < |a|n$. Suppose, conversely, that $|a| > 1/n$. Since $|\sin n\theta/\sin\theta|$ ranges between 0 and n for real θ, we can find a number $\varepsilon = e^{2i\theta}$, $0 < \theta < \pi$, such that

$$|1 + \varepsilon + \varepsilon^2 + \cdots + \varepsilon^{n-1}| = |(1 - \varepsilon^n)/(1 - \varepsilon)| = 1/|a|, \qquad |\varepsilon| = 1, \qquad \varepsilon \neq 1.$$

Next, determine w so that

$$w^{n-1}(1 + \varepsilon + \varepsilon^2 + \cdots + \varepsilon^{n-1}) = -1/a$$

and choose $z = \varepsilon w$. Then (*) will be fulfilled and hence $z + az^n = w + aw^n$.

(b) By part (a) we may assume that $|a| \leqslant 1/n$. The curve C_n may be parametrized by

$$x = \cos t + a\cos nt, \qquad y = \sin t + a\sin nt.$$

Since

$$(x')^2 + (y')^2 = 1 + a^2n^2 + 2an\cos(n-1)t \geqslant (1 - |a|n)^2,$$

we first consider the case $|a| < 1/n$. Now the curvature of C_n at the point $f(e^{it})$ is given by

$$K(t) = \frac{x'y'' - y'x''}{\{(x')^2 + (y')^2\}^{3/2}} = \frac{1 + a^2n^3 + an(n+1)\cos(n-1)t}{\{(x')^2 + (y')^2\}^{3/2}}.$$

The necessary and sufficient condition for C_n to be convex is that the curvature $K(t)$ does not change sign. But this is equivalent to the nonnegativity of

$$\min_{0 \leqslant t \leqslant 2\pi} \{1 + a^2n^3 + an(n+1)\cos(n-1)t\}$$
$$= 1 + a^2n^3 - |a|n(n+1) = (1 - n|a|)(1 - n^2|a|),$$

i.e., to the inequality $|a| \leqslant 1/n^2$. In the remaining case $a = \pm 1/n$, the curve C_n is the $(n-1)$-cusped epicycloid

$$x = \cos t \pm \frac{1}{n}\cos nt, \qquad y = \sin t \pm \frac{1}{n}\sin nt,$$

with cusps at the $n - 1$ points where $\cos(n-1)t = \mp 1$; accordingly, convexity does not hold in this case.

Editorial Comment. The proposer remarked that the condition that a be real is unnecessary, since, if, say, $a/|a| = e^{(n-1)i\alpha}$, then

$$e^{i\alpha}f(z) = (ze^{i\alpha}) + |a|(ze^{i\alpha})^n$$

Several readers noted that when $|a| \leqslant 1/n$, the curve C_n is actually starlike, i.e., $\arg f(e^{it})$ increases with t.

A necessary and sufficient condition for the convexity of the image of the unit circle $z = e^{it}(0 \leqslant t \leqslant 2\pi)$ under the function f is that $1 + \{\operatorname{Re} zf''(z)/f'(z)\} \geqslant 0$ for all z with $|z| = 1$. [Cf. Problem 108 of Part III of Pólya-Szegö]. This inequality is equivalent to the nonnegativity of the curvature as used above.

O. P. Lossers remarked that (a) occurs as Problem 239 in L. I. Volkovyskii, G. L. Lunts, and I. G. Aramanovich, *A Collection of Problems in Complex Analysis*, Pergamon Press, 1965. In that book it is mentioned that C_n is the trajectory of a point on a spoke of a circular wheel of radius $1/n$ which moves on the outside of a fixed circle of radius $(n-1)/n$, where the point on the spoke is at distance $|a|$ from the center of the wheel. This remark enables one to give a geometric argument for the result of (a).

Solved also by I. C. Bivens, L. R. King, and B. G. Klein (jointly), M. R. Gopal, H. O. Kim (Korea), P. L. Hon and T. S. Kin (jointly, Hong Kong), W. A. Newcomb, E. S. Rosenthal, V. Schindler, and the proposer.

Concurrent Conics

E 3172 [1986, 733]. *Proposed by Jordi Dou, Barcelona, Spain.*

Let A', B', C' be the feet of the altitudes from the vertices A, B, C respectively of a triangle ABC. Let H be its orthocenter and M be an arbitrary point of the plane. Prove that the conics $MABA'B'$, $MBCB'C'$, $MCAC'A'$, $MHAC'B'$, $MHAB'C'$ and $MHBC'A'$ have a point other than M in common.

Solution by O. P. Lossers, Eindhoven University of Technology, Eindhoven, The Netherlands. There are six lines in the configuration, viz. the three edges and the three altitudes of the triangle, which are pairwise intersecting in one of the given seven points A, B, C, A', B', C', H. Introduce homogeneous, projective coordinates (x_1, x_2, x_3) such that $A' = (1,0,0)$, $B' = (0,1,0)$, $C' = (0,0,1)$ and $H' = (1,1,1)$. Then $A = (-1,1,1)$, $B = (1,-1,1)$, $C = (1,1,-1)$, and the six lines have the equations $x_i \pm x_j = 0$, for $1 \leqslant i < j \leqslant 3$. Let $M = (\mu_1, \mu_2, \mu_3)$ be a point not collinear with any pair of these seven points, i.e., $\mu_i(\mu_j + \mu_k)(\mu_j - \mu_k) \neq 0$ for distinct i, j, k.

Instead of investigating the six conics, we consider an arbitrary cubic through the seven points and M. Its general form is

$$F(x_1, x_2, x_3) \equiv A_1 x_1(x_2^2 - x_3^2) + A_2 x_2(x_3^2 - x_1^2) + A_3 x_3(x_1^2 - x_2^2) = 0,$$

where the real numbers A_1, A_2, A_3 must be chosen so that $F(\mu_1, \mu_2, \mu_3) = 0$. Because $F(\mu_1^{-1}, \mu_2^{-1}, \mu_3^{-1}) = -(\mu_1, \mu_2, \mu_3)^{-2} F(\mu_1, \mu_2, \mu_3)$, we see that also $N = (\mu_1^{-1}, \mu_2^{-1}, \mu_3^{-1})$ is on this cubic. In particular, this is true if the cubic degenerates into one of the six lines with its corresponding conic, so M and N are both common points of the six conics. In the original situation, M and N are isogonally conjugate with respect to the triangle $A'B'C'$.

Also solved by H. Demir and C. Tezer (Turkey), E. Morgantini (Italy), and the proposer.

Zero-free Multiples

E 3196 [1987, 300]. *Proposed by I. Martin Isaacs, University of Wisconsin, Madison, and Dave Witte, University of California, Berkeley.*

Prove that every natural number n not divisible by 10 has a multiple whose decimal representation involves no 0's. (For example, 6144 is a multiple of 1024.)

Solution by the University of South Alabama Problem Group. If $(n, 10) = 1$, then $10^{\phi(n)} - 1$ suffices, since $10^{\phi(n)} \equiv 1 \bmod n$. Hence we may assume n is a multiple of 2 or 5 (but not both).

If $m \geqslant 1$, then there is an m-digit number $x = x_{m-1} \dots x_0$ such that $2^m | x$ and each x_i is 1 or 2. To construct x inductively let $x_0 = 2$, and if $x_{m-1} \dots x_0 = c2^m$, then let x_m be 1 if c is odd and 2 if c is even. This works because $x_m x_{m-1} \dots x_0 = (c + x_m 5^m)2^m$. A similar argument constructs an m-digit integer $y = y_{m-1} \dots y_0$ such that $5^m | y$ and each $y_i \in \{1, 2, 3, 4, 5\}$.

Now write $n = p^m a$, where p is 2 or 5 and $(a, 10) = 1$. A standard pigeonhole argument using partial sums of $10^0, 10^m, 10^{2m}, \dots$ shows that a has a multiple of the form $10^{sm} + 10^{(s+1)m} + \cdots + 10^{(s+r)m}$, which in turn yields $R = 1 + 10^m + \cdots + 10^{rm}$ as a multiple of a. Let $M = R \cdot (x_{m-1} \dots x_0)$ or $M = R \cdot (y_{m-1} \cdots y_0)$, depending on whether p is 2 or 5. Now M has digits in $\{1, 2, 3, 4, 5\}$ and is a multiple of both p^m and a, so M is a multiple of n.

Editorial comment. Eugene Levine pointed out that the digits of the zero-free multiple can be restricted to any set of five digits that form a complete residue system mod 5 and include at least one even digit.

Also solved by B. Gordon and J. M. Rojas (student), E. Levine, O. P. Lossers (The Netherlands), and the proposers. Partially solved by M. W. Ecker and R. A. Simon (Chile).

Birth of a Regular Hexagon

E 3198 [1987, 301]. *Proposed by I. J. Schoenberg, Madison, Wisconsin.*

Let $\Pi_x = (x_0, x_1, x_2, x_3, x_4, x_5) := (x_k)$, $0 \leqslant k \leqslant 5 (\mathrm{mod}\, 6)$ be an arbitrary closed skew hexagon in $\mathbb{R}^3$. Form a new hexagon $\Pi_y = (y_k)$ by

$$y_k = x_k + \frac{1}{2}(x_k - x_{k-3}) = \frac{3}{2}x_k - \frac{1}{2}x_{k-3}.$$

Finally, define a third hexagon $\Pi_w = (w_k)$ by

$$w_k = \frac{1}{3}(y_k + x_{k-1} + x_{k+1}).$$

Show that the hexagon Π_w lies in a plane π and is in π an affine image of a regular hexagon.

Solution by Jan van de Craats, Koninklijke Militaire Academie, Breda, The Netherlands. Since $w_k = \frac{1}{6}(2x_{k+1} + 3x_k + 2x_{k-1} - x_{k-3})$ for all k, we have $w_{k+3} - w_k = 2(w_{k+2} - w_{k+1})$ for all k. Any hexagon satisfying this property is

planar, since

(a) if three consecutive vertices coincide, then all vertices coincide,

(b) if three consecutive vertices are collinear, then all vertices are collinear, and

(c) if no three consecutive vertices are collinear, then all vertices lie in the plane determined by the first three.

Finally, let $\Pi_a = (a_0, \ldots, a_5)$ be an arbitrary regular hexagon. Then $\alpha(a_i) = w_i$ for $i = 0,1,2$ determines an affine map. Applying $a_{k+3} - a_k = 2(a_{k+2} - a_{k+1})$ for $k = 0,1,2$ yields $\alpha(a_i) = w_i$ for $i = 3,4,5$, which completes the proof.

Note that the proof generalizes to hexagons in affine spaces of dimension $\geqslant 2$ over an arbitrary commutative field K with $\mathrm{char}(K) \neq 2,3$. Note also that the degenerate cases may occur. If Π_x is the non-planar $((3,0,0), (-2,-2,1), (0,3,0), (1,-2,-2), (0,0,3), (-2,1,-2))$, then each vertex of Π_w is $(0,0,0)$; if $\Pi_x = ((0,0,0), (1,0,0), (0,1,0), (0,0,0), (1,0,0), (0,1,1))$, then $\Pi_w = ((2,2,2), (2,2,0), (2,2,-1), (2,2,0), (2,2,2), (2,2,3))$. Further material on affinely regular polygons appears in J. van de Craats and J. Simonis, Affinely regular polygons, *Nieuw Archief voor Wiskunde IV*, 4(1986) 225–240.

Editorial comment. There are several approaches to this problem. O. P. Lossers computed $v_k = w_k - (\Sigma x_i)/6 = (x_{k-1} + 2x_k + x_{k+1} - x_{k+2} - 2x_{k+3} - x_{k+4})/6$, from which $v_k = -v_{k+3}$ and $v_k + v_{k+2} + v_{k+4} = 0$ follow immediately and imply that Π_w is the plane affine image of a regular hexagon. The proposer began by expressing the x_k as the values of a fifth-degree polynomial at the sixth roots of unity.

Also solved by P. L. Hon (Hong Kong), O. P. Lossers (The Netherlands), W. A. Newcomb, V. C. and D. J. Williams, and the proposer.

A Superlogarithmic Density

E 3200 [1987, 301]. *Proposed by Paul Erdös, Hungarian Academy of Science, Budapest.*

Suppose $a_1, a_2, \ldots$ are real numbers such that $0 < a_1 < a_2 < \cdots, a_n \to \infty$, and $(a_1 + a_2 + \cdots + a_n)/a_n \to \infty$. If $N(x) = \Sigma_{a_n < x} 1$, prove that $N(x)/\log x \to \infty$ as $x \to \infty$.

Solution I by Bernt Lindström, University of Stockholm, Sweden. Put $S_n = a_1 + a_2 + \cdots + a_n$. Since $\log x < x - 1$ for $x > 1$, we have

$$\log S_n - \log S_{n-1} < (S_n - S_{n-1})/S_{n-1} < \varepsilon$$

when $n > N = N(\varepsilon)$, and $\log S_n - \log S_N < \varepsilon(n - N)$ when $n > N$. Hence $S_n < ce^{\varepsilon n}$ for $n > N$, where $c = c(\varepsilon)$. Given $x > ce^{\varepsilon N}$, determine n such that $ce^{\varepsilon n} < x \leqslant ce^{\varepsilon(n+1)}$. Then $n \geqslant (1/\varepsilon)\log(x/c) - 1$ and $N(x) \geqslant n$, since $a_1, a_2, \ldots, a_n < S_n < x$. Therefore $N(x)/\log x > 1/(2\varepsilon)$ when $x > X(\varepsilon)$. Hence $N(x)/\log x \to \infty$ as $x \to \infty$.

Solution II and generalization by Amram Meir, University of Alberta, Edmonton. Let $N(t)$ be nondecreasing on $[0, \infty)$, and let

$$A(x) = \frac{1}{x}\int_0^x t\, dN(t).$$

We claim

$$\underline{\lim}_{x\to\infty} A(x) \leqslant \underline{\lim}_{x\to\infty} \frac{N(x)}{\log x} \leqslant \overline{\lim}_{x\to\infty} \frac{N(x)}{\log x} \leqslant \overline{\lim}_{x\to\infty} A(x). \tag{1}$$

To prove (1), we have

$$\int_0^y \frac{A(x)}{x}\,dx = \int_0^y t\,dN(t) \int_t^y \frac{dx}{x^2} = \int_0^y dN(t)\left(1 - \frac{t}{y}\right) = N(y) - A(y). \tag{2}$$

If $A(x) > \lambda$ for all $x \geqslant x_0$, then $\int_0^y A(x)\,dx/x > \lambda \log(y/x_0)$ for $y > x_0$. By (2), this yields $N(y) > \lambda \log y + O(1)$ as $y \to \infty$. This proves the left-most inequality in (1). To prove the right-most inequality, we may assume that $\overline{\lim} A(x) < \infty$. If $A(x) < \mu$ for all $x > x_1$, then (2) implies $N(y) > \mu \log(y/x_1) + O(1)$ as $y \to \infty$, which completes the proof.

To obtain the proposed result as a special case, set $N(t) = k$ for $a_k \leqslant t < a_{k+1}$ and $k = 0, 1, \ldots$. Then $A(t) = (a_1 + a_2 + \cdots + a_k)/t$. The condition of the problem is that $A(a_k) \to \infty$. Since

$$A(a_{k+1}) - 1 = \frac{a_1 + a_2 + \cdots + a_k}{a_{k+1}}$$

$$\leqslant A(t) = \frac{a_1 + \cdots + a_k}{t} \leqslant \frac{a_1 + \cdots + a_k}{a_k} = A(a_k),$$

it follows that $A(t) \to \infty$. Using (1) we find that $N(x)/\log x \to \infty$.

Editorial Comment. Grahame Bennett noted that the result holds without the hypothesis that the sequence is increasing, as long as $a_n \to \infty$ and $(a_1 + a_2 + \cdots + a_n)/a_n \to \infty$.

Solved also by M. A. Arcones (Spain), G. Bennett, J. Ferrer (Spain), E. Hertz, K.-W. Lau (Hong Kong), O. P. Lossers (The Netherlands), L. E. Mattics, A. Riese, J. M. Rojas (student), J. H. Steelman, P. Tracy, and the proposer.

Polynomials Whose Values Are Divisible by *m* But Whose Coefficients Are Not

E 3203 [1987, 372] *Proposed by Hugh Montgomery, University of Michigan, Ann Arbor.*

Let m and n be positive integers. Show that the following two assertions are equivalent.

(i) There exists a polynomial

$$f(x) = \sum_{k=0}^{n} a_k x^k$$

with integral coefficients such that $\gcd(a_0, a_1, \ldots, a_n, m) = 1$ and $f(j)$ is divisible by m for all integers j.

(ii) $m|n!$.

Solution I by L. E. Mattics, University of South Alabama. Let $\Delta f(x) = f(x+1) - f(x)$. If $m|f(j)$ for all integers j, then $m|\Delta^i f(j)$ for all nonnegative integers i

and all integers j. Suppose p is prime and $p^a | m$ for some $a > 0$, and let q be the largest integer such that $\gcd(a_q, p) = 1$. Then p^a divides

$$\begin{aligned}\Delta^q f(0) &= \sum_{r=0}^{q} (-1)^{q-r}\binom{q}{r} f(r)\\ &= \sum_{k=0}^{n}\left(\sum_{r=0}^{q}(-1)^{q-r} r^k \binom{q}{r}\right) a_k\\ &= q!\left(S_q^n a_n + S_q^{n-1} a_{n-1} + \cdots + S_q^q a_q\right),\end{aligned}$$

where S_i^j are Stirling numbers of the second kind. We conclude that $p^a | q!$, because the second factor in the above product is relatively prime to p. Since $q \leqslant n$, $p^a | n!$. Hence $m | n!$.

To prove the converse, let $f(x) = x(x+1)\cdots(x+n-1)$. It is well known that $n! | f(i)$ for all integers i. Thus if $m | n!$, then $m | f(j)$ for all j.

Solution II by O. P. Lossers, Eindhoven University of Technology, The Netherlands. To see that (ii) implies (i), let $m | n!$. Then the polynomial $f(x) = \prod_{g=1}^{n}(x+g)$ has the desired properties, since it is monic and $|f(j)|$ for integer j is either 0 or $n!$ times a binomial coefficient.

To prove that (i) implies (ii), let $f(x)$ be as in assertion (i). Then $f(x)$ can be expressed as

$$f(x) = b_0 + \sum_{k=1}^{n} b_k \prod_{j=0}^{k-1}(x-j)$$

for a unique sequence of integers b_k, $k = 0, 1, \dots, n$. It can easily be seen that $\gcd(b_0, \dots b_n) = \gcd(a_0, \dots a_n)$. That $f(j) \equiv 0 \pmod m$ for the integers $j = 0, 1, \dots, n$ implies that

$$j! b_j \equiv 0 (\operatorname{mod} m) \quad \text{for } j = 0, 1, \dots, n. \qquad (*)$$

Since $\gcd(b_0, \dots, b_n, m) = \gcd(a_0, \dots, a_n, m) = 1$, there exist integers $C_0, \dots C_n$, D such that $\sum_{j=0}^{n} C_j b_j + Dm = 1$. Hence

$$\sum_{j=0}^{n} C_j b_j n! + Dmn! = n!. \qquad (**)$$

By $(*)$, the left-hand side of $(**)$ is a multiple of m, hence so is the right-hand side; that is, $m | n!$.

Editorial comment. The proposer notes that the equivalence of (i) and (ii) was established by G. Pólya in his article Über ganzwertige Polynome in algebraischen Zahlkörpern, *J. Reine Angew. Math.*, 149 (1919) 97–116.

Also solved by M. Chellali (France), C. B. Khare (India), J. H. Steelman, and the proposer.

A Formula for Power Sums

E 3204 [1987, 372]. *Proposed by Ira Gessel, Brandeis University.*

Let $S_r(n) = \sum_{i=0}^{n} i^r$. Prove that

$$\begin{aligned}S_{2r}(n) &= (2n+1)p_r(n(n+1)) \qquad (r \geqslant 1),\\ S_{2r+1}(n) &= q_r(n(n+1)) \qquad (r \geqslant 0),\end{aligned}$$

where the polynomials p_r and q_r are given by the generating functions

$$\frac{\sinh(x\sqrt{1+4t}/2)}{2\sqrt{1+4t}\sinh(x/2)} = \frac{1}{2} + \sum_{r=1}^{\infty} p_r(t)\frac{x^{2r}}{(2r)!},$$

$$\frac{\cosh(x\sqrt{1+4t}/2) - \cosh(x/2)}{2\sinh(x/2)} = \sum_{r=0}^{\infty} q_r(t)\frac{x^{2r+1}}{(2r+1)!}.$$

Cf. A. W. F. Edwards, A quick route to sums of powers, *Amer. Math. Monthly*, 93 (1986) 451–455.

Solution by the University of South Alabama Problem Group, Mobile, AL. Let $t = n(n+1)$ and replace x by ix in the generating functions above to obtain

$$\frac{\sin[(2n+1)x/2]}{2\sin(x/2)} = \frac{2n+1}{2} + \sum_{r=1}^{\infty}(-1)^r(2n+1)p_r(n(n+1))\frac{x^{2r}}{(2r)!},$$

$$\frac{\cos(x/2) - \cos[(2n+1)x/2]}{2\sin(x/2)} = \sum_{r=0}^{\infty}(-1)^r q_r(n(n+1))\frac{x^{2r+1}}{(2r+1)!}.$$

Replace the factors involving sin and cos by their exponential expressions; this enables the left sides of the two equations above to be rewritten as $(1/2) + \sum_{k=1}^{n}\cos kx$ and $\sum_{k=1}^{n}\sin kx$, respectively. Use the power series expansions of $\cos kx$ and $\sin kx$ in these expressions and interchange the order of summation. The formulas for S_r then appear by equating corresponding coefficients with the generating functions above.

Editorial comment. The proposer used the approach of forming the exponential generating function

$$F_n(x) = \Sigma_r S_r(n)x^r/r! = 1 + e^x + \cdots + e^{nx}$$
$$= [e^{(n+1/2)x} - e^{-x/2}]/[e^{x/2} - e^{-x/2}].$$

The identities and substitutions used above then yield the desired formulas for the coefficients $S_r(n)$. The proposer also noted that differentiation with respect to t yields $q_r'(t) = (2r+1)p_r(t)$.

Also solved by K. Dilcher (Canada), W. Janous (Austria), W. A. Newcomb, A. Pedersen (Denmark), H.-F. Yeung (Australia), and the proposer.

A Summation-Product Identity

E 3211*[1987, 457]. *Proposed by László Tóth, Satu Mare, Romania.*

Let $\omega(k)$ denote the number of distinct prime factors of the positive integer k and let (i, n) denote the greatest common divisor of the positive integers i and n. Express

$$\sum_{i=1}^{n} 2^{\omega((i,n))}$$

in terms of the prime factorization of n.

Solution by C. Georghiou, University of Patras, Greece. Denote the given sum by $a(n)$; we show that $a(n) = n\Pi_{p|n}(1 + 1/p)$. Grouping the terms of the sum according to the value of (i, n), we obtain the convolution

$$a(n) = \sum_{d|n} 2^{\omega(d)} \sum_{\substack{1 \leqslant k \leqslant n/d \\ (k, n/d)=1}} 1 = \sum_{d|n} 2^{\omega(d)}\phi(n/d),$$

where ϕ is the totient function. Since $2^{\omega(n)}$ and $\phi(n)$ are multiplicative arithmetic functions, $a(n)$ is also multiplicative. (Cf. Niven and Zuckerman, *An Introduction to the Theory of Numbers*, §4.4.) Therefore, it suffices to evaluate $a(p^m)$, which is $p^m(1 + 1/p)$.

Editorial comment. Walther Janous and Hubert Kiechle both commented that replacing 2 by an arbitrary complex number z in the definition of a leads to $a(n) = n\Pi_{p|n}(1 + (z - 1)/p)$.

Also solved by the proposer and 21 other readers.

ADVANCED PROBLEMS

6586. *Proposed by Charles S. Allen, Drury College, Springfield, MO.*

In an infinite-dimensional normed linear space does there exist a convex set which does not contain a ray but whose closure does contain a ray?

6587. *Proposed by Ilan Vardi, Stanford University, CA.*

(a) The Kloosterman sum $S(m, n, c)$ is defined for positive integers m, n, c by

$$S(m, n, c) = \sum_{k=1}^{c}{}^{*} e^{2\pi i(mk + n\bar{k})/c},$$

where the asterisk indicates that the summation is restricted to integers k relatively prime to c and $\bar{k}$ is defined by $k\bar{k} \equiv 1 (\operatorname{mod} c)$. If c is an odd prime, compute the determinant of the $c - 1$ by $c - 1$ matrix such that the element in the jth row and kth column is $S(j, k, c)$ for $1 \leqslant j, k \leqslant c - 1$.

(b) Generalize Part (a) to composite c and the more general Kloosterman sum

$$S(m, n, \chi, c) = \sum_{k=1}^{c}{}^{*} \chi(k) e^{2\pi i(mk + n\bar{k})/c},$$

where χ is a Dirichlet character modulo c, i.e., a map from the integers to the cofnplex numbers such that

$$\chi(1) = 1, \chi(mn) = \chi(m)\chi(n), \chi(m + c) = \chi(m), \chi(m) = 0 \quad \text{if} \quad (m, c) > 1$$

for all m, n.

6588. *Proposed by Walther Janous, Ursulinengymnasium, Innsbruck, Austria.*

As in E 3246 put $\psi(n) = n\phi(n)$, where ϕ denotes the arithmetic function of Euler. For positive real x let $F(x)$ denote the number of positive integers n for which $\psi(n) \leqslant x^2$. Prove that $\lim_{x\to\infty} F(x)/x$ exists and find its value.

REVIEWS

EDITED BY JOSEPH KONHAUSER
Department of Mathematics, Macalester College, St. Paul, MN 55105

Applied Differential Geometry. By William L. Burke, Cambridge University Press, 1985. xvii + 414 pp.

DEANE YANG
Department of Mathematics, Rice University, Houston, TX 77251

Ideally, differential forms should be introduced as a natural part of multivariable calculus. One book that does this nicely is [E]. I hope that in time virtually all advanced calculus texts will adopt this approach. For the moment most of them do not, creating the need for books such as *Applied Differential Geometry*, which introduces the foundations of modern differential geometry, including manifolds, tangent and cotangent bundles, and differential forms. Most such texts, even those directed towards non-mathematicians, introduce differential forms by first discussing exterior algebra and the definition of a manifold and its tangent bundle. I dislike this approach, because the student is forced to absorb a lot of abstract material without any idea what it is good for. I would prefer a more heuristic approach at the start, something like the following:

I will assume that you are familiar with the multivariable calculus. First, I will introduce differential forms over a domain $D \subset \mathbb{R}^n$. Later, when we understand everything a little better, we will define what a manifold is and extend the ideas to the more general setting.

You have already encountered differential forms back in basic integral calculus. Recall that the integrand of a single variable integral is not simply a function $f(x)$ but $f(x)\,dx$, where the symbol dx indicates the variable of integration. The symbol dx is often called a differential, but no precise definition of what it is is ever given in an elementary course. One simply learns how to use it to change variables in the integral. There is only one rule: if y is a function of x, then

$$dy = \frac{dy}{dx}\,dx.$$

In fact, differentials can be given a precise mathematical definition, which we will indicate later. For now, however, let's just view them as symbols which obey certain rules.

Motivated by the notation of a differential as described above, we extend them to the general notion of a **1-form** on a domain $D \subset \mathbb{R}^n$ to be an expression of the form

$$f_1(x)\,dx^1 + \cdots + f_n(x)\,dx^n,$$

where $x = (x^1, \ldots, x^n)$ are the standard coordinates of $\mathbb{R}^n$ and $f_1, \ldots, f_n$ are smooth functions on D. Given 1-forms $\omega = f_1\,dx^1 + \cdots + f_n\,dx^n$ and $\theta = g_1\,dx^1 + \cdots + g_n\,dx^n$ and a smooth function $\varphi(x)$, we can add ω and θ to get

$$\omega + \theta = (f_1 + g_1)\,dx^1 + \cdots + (f_n + g_n)\,dx^n,$$

and multiply φ times ω to get

$$\varphi\omega = \varphi f_1\, dx^1 + \cdots + \varphi f_n\, dx^n.$$

Given a function $f(x)$, we define the **differential of f** to be the 1-form

$$df = \frac{\partial f}{\partial x^1}\, dx^1 + \cdots + \frac{\partial f}{\partial x^n}\, dx^n. \tag{1}$$

Observe that the chain rule implies that this definition is consistent under a change of independent variables, since if we set $x^i = x^i(u)$, where $u = (u^1, \ldots, u^n)$, and view dx^i as the differential of the function x^i, we obtain

$$\begin{aligned} df &= \frac{\partial f}{\partial x^1}\, dx^1 + \cdots + \frac{\partial f}{\partial x^n}\, dx^n \\ &= \sum_{i,j=1}^{n} \frac{\partial f}{\partial x^i}\frac{\partial x^i}{\partial u^j}\, du^j \\ &= \frac{\partial f}{\partial u^1}\, du^1 + \cdots + \frac{\partial f}{\partial u^n}\, du^n. \end{aligned}$$

As observed before, the integrand of a 1-variable integral is a 1-form on an interval in $\mathbb{R}$. The rules above and the change of variables theorem conspire to keep the value of the integral invariant under a change of independent variable. More generally, the integrand of a line integral along a curve in $\mathbb{R}^n$ is a 1-form on $\mathbb{R}^n$. Given the 1-form ω above and any curve $\gamma\colon [a, b] \to \mathbb{R}^n$, define the **integral of ω along γ** to be

$$\int_\gamma \omega = \int_a^b f_1(\gamma(t))\frac{dx^1}{dt} + \cdots + f_n(\gamma(t))\frac{dx^n}{dt}\, dt,$$

where $\gamma(t) = (x^1(t), \ldots, x^n(t))$. Observe that this definition is consistent with the definition of a 1-form on $\mathbb{R}$, where $x^1, \ldots, x^n$ are no longer viewed as independent variables but as functions of the new independent variable t. The value of this integral is invariant not only under changing the coordinate t but also under changing the coordinates $x^1, \ldots, x^n$.

In the traditional view of things, the ease of changing the variables is lost for a multiple integral. The differentials which still appear in the integrand no longer seem to help in computing the formula, which involves the determinant of the Jacobian of the old coordinates viewed as functions of the new. By introducing the notions of p-forms, p a positive integer, and of exterior multiplication, we can change variables of a multiple integral very easily without memorizing any complicated formulas. I make no claim that what follows is an obvious way to proceed. On the other hand, I do believe that it is better to introduce the basics of exterior algebra in a familiar context, where the student sees its utility, rather than in an abstract setting.

The first hint of what to do is that the integrand of a multiple integral is always of the form $f(x)\, dx^1 \cdots dx^n$, where each dx^i appears once and only once. Let's change variables, setting $x^i = x^i(u)$, where $u^1, \ldots, u^n$ are the new coordinates.

Proceeding formally, we replace each dx^i by

$$\frac{\partial x^i}{\partial u^1}du^1 + \cdots + \frac{\partial x^i}{\partial u^n}du^n,$$

and multiply them all together. Yet somehow, when we do this, we should end up with

$$\det\left[\frac{\partial x^i}{\partial u^j}\right]du^1 \cdots du^n.$$

First, given a positive integer k, we define a ***k*-form** on D to be an expression of the form

$$\sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n} a_{i_1 i_2 \cdots i_k}\, dx^{i_1}\, dx^{i_2} \cdots dx^{i_k},$$

where $a_{i_1 \cdots i_k}$ is a smooth function of $x \in D$. Such an expression looks just like a homogeneous polynomial in the "variables" $dx^1, \dots, dx^n$ of degree k, except that each variable appears only once in each term. Observe that due to this restriction, there are no k-forms for $k > n$. Also, an n-form is always of the form

$$a(x)\, dx^1 \cdots dx^n.$$

For convenience we shall also view functions as 0-forms.

Next, we want to say how to multiply differential forms together. The usual polynomial multiplication doesn't give us what we want. Define a new type of multiplication, called **exterior multiplication**, which is still associative and distributive (with respect to addition) but is no longer commutative. Given a p-form and a q-form, obtain a $(p+q)$-form by multiplying formally and by rearranging the order of the factors in each term using the following rules: Given a 0-form f and 1-forms θ, ω,

(1) 0-forms commute with everything;

(2) $\theta\omega = -\omega\theta$.

Observe that (2) implies that $\theta^2 = \theta \cdot \theta = -\theta \cdot \theta = 0$. In particular, terms which contain more than one copy of dx^i, for some i, vanish. The rules (1) and (2) imply the more general property that given a p-form Ω and a q-form Θ,

$$\Omega\Theta = (-1)^{pq}\Theta\Omega.$$

The reason that differential forms are so useful for integration is given by the following:

LEMMA. *Given functions $f^1, \dots, f^n$ on a domain $D \subset \mathbb{R}^n$,*

$$df^1\, df^2 \cdots df^n = \det\left[\frac{\partial f^i}{\partial x^j}\right] dx^1\, dx^2 \cdots dx^n.$$

Without computing the left-hand side in detail, it is clear that

$$df^1 \cdots df^n = \left(\text{polynomial in } \partial f^i/\partial x^j\right) dx^1 \cdots dx^n.$$

It is not hard to see that (2) implies that the polynomial is an alternating multilinear function of the rows of the Jacobian matrix and that it is equal to 1 when the Jacobian is the identity matrix. It follows that it is equal to the determinant.

Therefore, if we are given an integral

$$\int f(y)\, dy^1 \cdots dy^n$$

and want to change coordinates from $(y^1, \ldots, y^n)$ to new coordinates $(x^1, \ldots, x^n)$, we obtain the correct integrand by simply viewing $y^1, \ldots, y^n$ as functions of x and applying the rules of exterior multiplication.

There is, however, one not-so-slight catch. Since the change of variables formula uses the determinant of the Jacobian and not its absolute value, the integrals in this paper are *oriented integrals over oriented domains*. In practice, what this means is that when changing variables, one should always order the new coordinates so that the determinant of the Jacobian is positive.

So far, it appears that the only forms on $\mathbb{R}^n$ of real interest are 0-, 1-, and n-forms. On the other hand, the integrand of a surface integral is a 2-form on $\mathbb{R}^3$ and in general, the integrand of an integral over a k-dimensional "surface" in $\mathbb{R}^n$ is a differential k-form.

After this, we should define the exterior derivative of a differential form and observe that, like the differential of a function, it is well-defined and independent of the choice of coordinates. Also, in $\mathbb{R}^3$, the exterior derivative of 0-, 1-, and 2-forms encodes the familiar formulas for gradient, curl, and divergence. Finally, one recalls Stoke's theorem and observes that it has a simple and elegant formulation using differential forms.

By now, we hope that the student is convinced of the usefulness of this notation. It is time to give a precise mathematical definition of differential forms and give it a geometric interpretation.

Begin with 1-forms on $D \subset \mathbb{R}^n$. The idea is to view $dx^1, \ldots, dx^n$ as basis vectors of some vector space, call it V. A 1-form $\omega = f_1\, dx^1 + \cdots + f_n\, dx^n$ is then a map ω: $D \to V$. We prefer to set $T^*D = D \times V$ and view ω as a map

$$\begin{aligned} \omega \colon D &\to T^*D \\ x &\mapsto \big(x, f_1(x)\, dx^1 + \cdots + f_n(x)\, dx^n\big). \end{aligned}$$

Now what happens when we change coordinates? Let the new coordinates be $u = (u^1, \ldots, u^n)$ and suppose that $x^i = x^i(u)$. Then for each $x \in D$, we get a new basis of V, denoted $du^1, \ldots, du^n$, which is related to the old basis $dx^1, \ldots, dx^n$ by the equation

$$dx^i = \frac{\partial x^i}{\partial u^1}\, du^1 + \cdots + \frac{\partial x^i}{\partial u^n}\, du^n.$$

Thus, for each choice of coordinates on D, we get a new basis of $\{x\} \times V$. Using this observation, we want to identify T^*D a little better.

To do this, we look at a different vector space. Let D be as before, let T be another n-dimensional vector space with basis $\tau_1, \ldots, \tau_n$, and let $T_*D = D \times T$. Abstractly, we want to view $\{x\} \times T$ as the vector space of all possible tangent vectors to curves passing through x. The coordinates $x^1, \ldots, x^n$ induce a natural basis of T by specifying that τ_i is the tangent vector to the curve obtained by fixing x^j, $j \neq i$, and setting $x^i = t$.

Let $\gamma: [-\varepsilon, \varepsilon] \to D$ be a differentiable curve. Then the derivative of γ determines a map

$$\gamma_*: [-\varepsilon, \varepsilon] \to T_*D$$
$$t \mapsto (\gamma(t), \gamma'(t)),$$

where the tangent vector to γ at $\gamma(t)$ is

$$\gamma'(t) = \frac{dx^1}{dt}\tau_1 + \cdots + \frac{dx^n}{dt}\tau_n.$$

As before, a new set of coordinates $u^1, \ldots, u^n$ induce a new basis of T, call it $\mu_1, \ldots, \mu_n$. The two bases are related by the chain rule:

$$\begin{aligned}\gamma'(t) &= \frac{dx^1}{dt}\tau_1 + \cdots + \frac{dx^n}{dt}\tau_n \\ &= \sum_{i=1}^{n} \frac{\partial x^1}{\partial u^i}\frac{du^i}{dt}\tau_1 + \cdots + \frac{\partial x^n}{\partial u^i}\frac{du^i}{dt}\tau_n \\ &= \sum_{j=1}^{n} \frac{du^1}{dt}\frac{\partial x^j}{\partial u^1}\tau_j + \cdots + \frac{du^n}{dt}\frac{\partial x^j}{\partial u^n}\tau_j.\end{aligned}$$

From this we see that

$$\mu_i = \sum_{j=1}^{n} \frac{\partial x^j}{\partial u^i}\tau_j.$$

The set T_*D is called the **tangent bundle of D** and an element $(x, v) \in T_*D$ consists of a point x and a tangent or velocity vector sitting at x.

We can now say what V is. Note that the change of basis formula relating $dx^1, \ldots, dx^n$ to $du^1, \ldots, du^n$ is exactly the same formula for the change of basis for the dual bases to $\tau_1, \ldots, \tau_n$ and $\mu_1, \ldots, \mu_n$. In other words, using the coordinates $x^1, \ldots, x^n$, we can identify $V \cong T^*$ by viewing dx^i as the linear function on T satisfying $dx^i(\tau_j) = \delta_{ij}$. The formulas above then imply that this identification is fixed under changes of coordinates. Therefore, the space of 1-forms is, in some sense, naturally dual to the space of tangent vectors. For this reason, T^*D is called the **cotangent bundle of D**.

What about k-forms in general? They are, roughly speaking, dual to the space of k-planes in T_*D, but we will not try to be more precise about this.

Observe that the object with natural geometric meaning is the tangent bundle, since it is easy to draw pictures of velocity vectors and sums of such vectors. It is not so easy to draw corresponding pictures of 1-forms. Differential forms should not be viewed as natural geometric objects, meaning that one should not try too hard to visualize what they are. Rather they are natural *linear algebraic* objects obtained from the tangent space. For reasons we have no space to go into, it turns out that differential forms are much easier to manipulate than vector fields and therein lies their utility.

By now the idea of studying calculus using a coordinate-free approach is well established. The definitions of a manifold, its tangent and cotangent bundles, differential forms, exterior derivative, and integration now follow easily. Of course,

a quick course on exterior algebra should be included somewhere in all of this, now that the student is better motivated to learn it.

And what of the book *Applied Differential Geometry*? The emphasis is on explaining the ideas and the applications of the mathematics and not on a rigorous definition-theorem-proof exposition. I like that. I do, however, find the book too vague and sketchy for my taste. Pictures and heuristic discussions are essential, but they cannot replace precise definitions and theorems. The book does not contain much mathematics or physics. The author's mistake, I think, was to include material on too many different topics. He would have done better to cover only a few of them in more depth. He mentions that he taught an entire graduate course on electromagnetism using differential forms. This should be a much more effective way of introducing the notion to physics students.

The author draws a lot of pictures of what the mathematical notation represents. Although many of them are very instructive, most of the explanations are inadequate. Also, he goes too far, I think, in trying to draw a picture of everything. It is not hard to draw pictures of curves, surfaces, tangent vectors, and tangent planes. Adding vectors pictorially is easy. The same cannot be said for cotangent vectors or differential forms. Nevertheless, the author includes drawings which "explain" what differential forms are, how to add two k-forms, how the Hodge star operator works, etc.

As a professional differential geometer, I use differential forms all the time and must be able to visualize what my calculations and formulas mean. This does *not* mean visualizing differential forms themselves. The formulas and calculations are meant to represent geometric assertions involving manifolds and their tangent bundles. That is what I draw pictures of. What is important to me is the ability to translate with ease between equations and computations involving differential forms and the equivalent formulation using tangent vectors and submanifolds.

Books such as *Applied Differential Geometry* or **[F]** concentrate on the basics of differential forms and give a few simple applications in physics. Differential forms, however, are a powerful tool in pure mathematics, particularly geometry and topology. Elementary expositions can be found in **[GP]** and **[O]**. For advanced applications in topology, try **[BT]**; in differential geometry, **[BCG]** and **[C]**; and in algebraic geometry, **[GH]**.

REFERENCES

[BT] R. Bott and L. Tu, Differential Forms in Algebraic Topology, Springer-Verlag, New York, 1982.

[BCG] R. Bryant, S. S. Chern, P. Griffiths, et al. Essays in Exterior Differential Systems, Springer-Verlag, New York, to appear.

[C] E. Cartan, Les systèmes différentiels extérieurs et leurs applications géométriques, Hermann, Paris, 1963.

[E] H. M. Edwards, Advanced Calculus, Houghton Mifflin, Boston, 1969.

[F] H. Flanders, Differential Forms with Applications to the Physical Sciences, Academic Press, New York, 1963.

[G] P. Griffiths, On Cartan's method of Lie groups and moving frames as applied to existence and uniqueness questions in differential geometry, *Duke J. Math.*, 41 (1974) 775–814.

[GH] P. Griffiths, and J. Harris, Algebraic geometry and local differential geometry, *Ann. Scient. Ec. Norm. Sup.*, 12 (1979) 355–432.

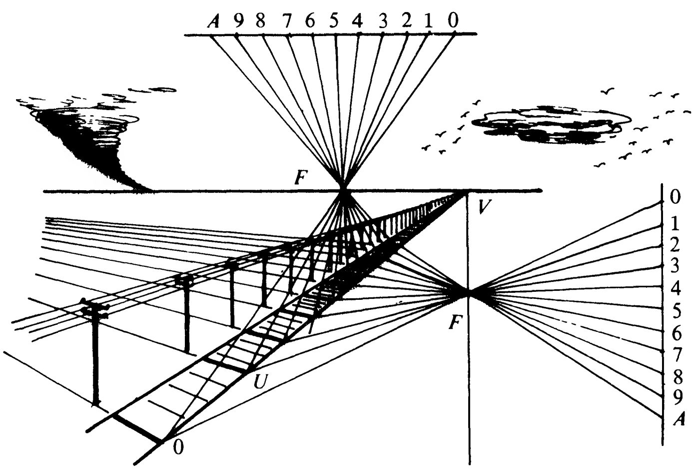
[GP] V. Guillemin, and A. Pollack, Differential Topology, Prentice Hall, Englewood Cliffs, N.J., 1974.

[MTW] C. Misner, K. Thorne, and J. Wheeler, Gravitation, W. H. Freeman and Co., San Francisco, 1973.
[O] B. O'Neill, Elementary Differential Geometry, Academic Press, New York, 1966.
[S] M. Spivak, A Comprehensive Introduction to Differential Geometry, vols 3, 4, 5, Publish or Perish, Inc., Wilmington, Del., 1979.

A Topological Picturebook. By George K. Francis. Springer-Verlag, New York, 1987. xv + 194 pp.

JEFFREY R. WEEKS
Mathematics and Computer Science Department, Ithaca College, Ithaca, NY 14850

Perspective drawing is a great introduction to projective geometry. The experience gained in tackling the concrete problems of perspective is not only valuable in itself, but leads students to a deeper appreciation of the abstract mathematics which follows. While some problems of perspective are easy enough to build confidence in the weaker students, others are quite challenging. Consider, for example, the problem of drawing railroad tracks stretching out across a prairie to the horizon. Drawing the rails is easy enough, but where is the vanishing point for the ties? And how should the ties be spaced? FIGURE 1, from the book under review, illustrates a method for spacing the ties. Pick any convenient focus point F distinct from the vanishing point V and draw the line FV. (Figure 1 illustrates two possible choices of F. In one case the line FV happens to coincide with the horizon. In the other it happens to be perpendicular to the horizon.) Pick an origin O and a unit point U on one of the rails. The distance OU is the unit relative to which other distances along the rail will be measured. Draw any line r parallel to FV, and project O and U through F to r. The line r now serves as a perspective ruler for measuring distances along the rail. In particular, a sequence of equally spaced points on the ruler projects back through the focus F to a correct perspective view of equally spaced



Horizon and Zenith

FIG. 1.

points on the rail. This solves the problem of how to space the railroad ties. The justification of the method starts the student on the road to an understanding and appreciation of the abstract and powerful ideas of projective geometry. (The method's validity follows as a corollary of the projective invariance of the cross ratio. Yet proofs involving no mathematics beyond high school geometry are also possible. The reader is challenged to find one.)

The preceding example illustrates how the concrete problems of perspective drawing prepare a student for the abstractions of projective geometry. The generalization of this pedagogical principle pervades every aspect of the book under review: specifically, *A Topological Picturebook* is based on the principle that an understanding and appreciation of concrete examples enhances one's understanding and appreciation of the abstract theorems they illustrate. The book's very existence embodies the principle. Its stated purposes are "to encourage mathematicians to illustrate their work, and to help artists to understand the abstract ideas expressed by such drawings." In other words, it is a book which presents examples and encourages others to do so. Its internal structure also reflects the principle. The author skillfully presents examples of a technique *before* discussing the technique explicitly. For example, the book opens with a delightful discussion of how to draw a saddle in a box and in a drum (the box and the drum are not part of the final drawings—they serve only to bound the region of space to which the drawing of the saddle is confined). Only after presenting these examples does the author explicitly remark that drawing a surface clipped to the boundaries of a familiar region (in this case a cube or cylinder) is more effective than drawing it in reference to coordinate axes, because in everyday life people orient themselves relative to boundaries such as the walls of a room, not relative to thin rods with tick marks. The reader readily appreciates the author's remark because he has already seen the method in action.

This same principle, that an appreciation of nontrivial, concrete examples should precede the introduction of more general, abstract theorems, deserves wider application in the classroom, particularly in our upper-level courses. Students should begin a new topic by tackling some problems and working with some examples. The problems and examples must be diverse and nontrivial to prepare the students for what follows. For instance, students who have seen only the integers and the integers mod n will not appreciate the need for an abstract definition of a group. Once the students are comfortable with the examples, then it's time for a theorem. As far as possible let the students themselves search for the correct statement. How might we prove the theorem? Again, let the students discuss their own ideas. There is no need to say whether their original guess for the statement of the theorem was correct or not; let them work it out themselves. Their extensive experience with examples and problems will help them sort the grain from the chaff. The final step, once the theorem and the outline of a proof are understood, is rigor. Details are important: they come last because only a student who is comfortable with the big picture can fully appreciate them.

As an example of the above idea, consider the teaching of topology. When I was an undergraduate my first topology course contained nothing but point-set topology. It was interesting stuff, but it dwelt upon transfinite sets, infinite products, and so on, and never got as far as what I considered *real* topology. My conception of real topology was pretty dim, but I felt it should have something to do with twisted surfaces and other interesting shapes. (I now know that what I was looking for is

called geometric topology, but unfortunately it wasn't part of the curriculum.) My second course in topology began with a tedious exercise in subdividing simplices, followed by the definition—in one fell swoop—of an infinite series of groups. In retrospect it's clear that the instructor was defining the homology groups, but at the time I could find no meaning in it at all and I dropped the course. The problem was not that the homology groups are intrinsically uninteresting or that the instructor's presentation was unclear; the problem was that I was intellectually unprepared to appreciate them. I would have learned much more had I begun with some concrete examples and problems. I would have loved to have gotten my hands on some interesting manifolds. Eventually the subtler notions would arise. For example, the surface obtained by identifying opposite sides of a square is somehow the same as the surface obtained by identifying opposite sides of a hexagon. In what sense are they "the same"? Here's a great place to talk about homeomorphism! Furthermore, if different constructions can yield the same surface, how can we ever be really sure that two surfaces or manifolds are different? We could look at the properties of loops contained in the manifolds. On a sphere every loop can be shrunk to a point, but on a torus some cannot; this observation provides an informal proof that the sphere and the torus are not homeomorphic. Students who have internalized these ideas—by seeing examples and working problems—are ready to appreciate the homology and homotopy groups.

I should emphasize that I do *not* advocate less abstraction or less rigor. To the contrary, I'm disturbed by the watering down of our upper-level courses. I'm suggesting that more hands-on experience with examples and problems prepares students for the abstraction and the rigor.

I'd now like to comment on *A Topological Picturebook* itself.

The organization of the book is imaginative and effective. The drawing lessons are nominally to be found in the first three chapters ("Descriptive Topology," "Methods and Media," and "Pictures in Perspective"). The five remaining chapters are "picture stories": each contains a series of pictures, accompanied by some explanatory text, which leads the reader to an understanding of some especially delightful examples from twentieth-century topology, such as sphere eversions or the fibration of the figure-eight knot complement. The division between the methods chapters and the picture story chapters is not strict. The methods chapters use a picture story format in which each section is a mini-picture-story illustrating the technical point under consideration. And the picture story chapters are sprinkled with hints on making the drawings, each appearing in a natural context.

The book's two purposes, as previously stated, are "to encourage mathematicians to illustrate their work and to help artists to understand the abstract ideas expressed by such drawings." The book succeeds nicely at its first purpose. By carefully interweaving the drawing lessons with concrete and interesting examples, it lets mathematicians learn about drawing while reading about math. Unfortunately the book fails at its second purpose. It is too difficult and too incomplete for a nonmathematician. For example, Boy's surface plays a central role in the chapter on Shadows from Higher Dimension, yet the reader is referred elsewhere for a picture of Boy's surface and an introduction to its topology.

The book achieves a third, unstated purpose which is every bit as valuable as the first. The picture stories present a few of the most appealing examples in topology,

and are well worth reading for their own sake. (The incompleteness mentioned in the preceding paragraph will occasionally annoy even a professional mathematician, but fortunately the author did his homework and provided complete bibliographic references.) All too often the examples which motivate an author's thinking get lost amongst the formalism when he goes to publish his work. It's nice to see these examples taking their place in the literature, for our pleasure as well as our understanding of topology.

In appreciation...

At this time of the year we would like to thank those who have refereed papers for the Monthly during the past year. Without their diligent efforts we could not have functioned.

—**The Editors**

Harvey L. Abbott, Arnold Adelberg, S. G. Akl, Michael Albertson, Felix R. Albrecht, Leo J. Alex, J. Ralph Alexander, C. D. Aliprantis, Steven Althoen, John Anderson, William N. Anderson, Jr., George E. Andrews, Tom Apostol, Kenneth Appell, Richard M. Aron, Richard A. Askey, Raymond Ayoub, Sara Baase, Gennady Bachman, Steven B. Bank, Edward J. Barbeau, Jr., M. F. Barnsley, Michael Barr, Peter W. Bates, Donna Beers, Lowell W. Beineke, Howard E. Bell, Edward A. Bender, Harold Benzinger, I. D. Berg, Marc A. Berger, Earl R. Berkson, Bruce Berndt, Patrick Billingsley, Richard L. Bishop, William A. Blankinship, Harvey Blau, Lenore Blum, Robert Blumenthal, Ralph P. Boas, Kenneth Bogart, Richard C. Bollinger, Peter Borwein, M. Boshernitzan, W. E. Boyce, David Boyd, Robert S. Boyer, Walter Brady, Michael Branton, Priscilla Bremser, William E. Briggs, John Brillhart, Duane M. Broline, Richard Brualdi, Hugh D. Brunk, James D. Buckholtz, R. P. Burn, Stephen A. Burr, Bob Burton, Richard D. Byrd, Bryan E. Cain, Rodney Canfield, Michael A. Carchidi, Francis Carroll, Robert W. Carroll, Frank S. Cater, Gulbank D. Chakerian, S. D. Chatterji, Barry A. Cipra, George Cobb, Ernest J. Cockayne, John Coffey, Daniel A. Cohen, David Cohen, Graeme L. Cohen, Paul M. Cohn, Susan Colley, Karen Collins, W. Wistar Comfort, Roger Cooke, William P. Cooke, Curtis Cooper, David Cox, John Cremona, Robert J. Currier, Karl David, Everett C. Dade, J. M. Anthony Danby, John P. D'Angelo, Chandler Davis, Mahlon M. Day, Robert Devaney, Frank Deutsch, Persi Diaconis, Ralph L. Disney, Michael J. Dixon, Michael Dollinger, David Dorman, Patrick Dowling, Michael P. Drazin, Underwood Dudley, Paul Edelman, Herbert Edelsbrunner, Larry Eggan, Ronald J. Evans, John A. Ewell, Charles vanden Eynden, Kenneth J. Falconer, Hanafi K. Farahat, Leslie Jane Federer, Michael A. Filaseta, Nathan J. Fine, Steve Fisk, David C. Fisher, Harley Flanders, Gerald B. Folland, C. K. Fong, Paul Fong, David Ford, Peter Fowler, John Franke, George K. Francis, Greg Frederickson, David Freedman, Gerd H. Fricke, Shmuel Friedland, Wolfgang H. J. Fuchs, Joseph A. Gallian, Fred Galvin, Marianne Gardner, Adriano M. Garsia, Frank Garvan, Susan Geller, Murray Gerstenhaber, Ira Gessel, Robert Gethner, Richard A. Gibbs, Peter Gibson, Janice Gifford, Robert Gilman, Clark Givens, Andrew Gleason, James Glimm, Henry Glover, David Gluck, Herman Gluck, Donald Y. Goldberg, R. Goldman, Daniel A. Goldston, Irving J. Good, Judith Grabiner, Ronald L. Graham, Andrew Granville, Alfred Gray, John W. Gray, Daniel R. Grayson, Ralph P. Grimaldi, Stanley Grossman, Emil Grosswald, Branko Grünbaum, Heinrich Guggenheimer, William H. Gustafson, Peter Hagis, Molly Hahn, Mary-Elizabeth Hamstrom, Shih-Ping Han, James Hannan, David Harbater, Ferdinand Haring, Theodore E. Harris, Morton E. Harris, Robert E. Hartwig, David Hayes, John Hayes, Peter Hayslett, Henry E. Heatherly, Robert Hecht-Nielson, Allan G. Heinicke, Lester L. Helms, Jim Henle, Douglas A. Hensley, Ralph Henstock, Jack Herriot, Nigel Higson, Theodore Hill, David Lee Hilliker, Peter Hilton, A. Howard M. Hoare, Arthur Hobbs, Jack Hodges, Richard Holley, Joseph Horowitz, Fredric T. Howard, Tony Hughes, James E. Humphreys, F. J. Hwang, Yasubiko Ikebe, Ron Irving, I. M. Isaacs, John R. Isbell, Mourad E. H. Ismail, Steve Janke, Thomas Jech, Richard Jerrard, Carl G. Jockusch, Robert H. Johnson, Wells Johnson, Elgin H. Johnston, Kevin W. Kadell, William Kantor, Talbot Katz, Louis Kauffman, Robert P. Kaufman, Nicholas Kazarinoff, Edward Kerner, Jeanne Kerr, Riaz Khan, Kent Kidman, Clark H. Kimberling, David Klarner, Lee Klinger, William Knight, Murray S. Klamkin, Victor L. Klee, Daniel J. Kleitman, Frank B. Knight, Ron Knill, Marvin I. Knopp, David Knuth, Fred Kochman, Marc

Konvisser, Gregers Krabbe, Mark Krusemeyer, Kenneth Kunen, Jeanne LaDuke, John Ladwig, Jeffrey C. Lagarias, T. Y. Lam, Timothy Lance, Martin Landau, Richard Lane, Ray G. Langebartel, David D. Lantz, M. Paul Latiolais, Richard Laver, Solomon Leader, Eugene Lee, Sin-Min Lee, Anne Leggett, Derrick H. Lehmer, Emma Lehmer, Joseph Lehner, Bill Lenhart, Russell V. Lenth, Maita Levine, Jonathan Lewin, Joseph J. Leing, C. L. Liu, Robin H. Lock, Paul E. Long, N. J. Lord, Edith Luchins, Frederick W. Luttmann, Jr., Wilhelmus A. J. Luxemburg, Saunders Mac Lane, J. J. Malone, Marvin Marcus, Matthew J. Markert, Thomas Markham, W. S. Martindale, T. McCready, Leon R. McCulloh, Kevin McCurley, O. Carruth McGehee, Lynn McLinden, Elizabeth McMahon, David G. Mead, Gary H. Meisters, Paul Melvin, Bert Mendelson, Ron Menton, Frederic T. Metcalf, Nicholas C. Metropolis, Christian Meyer, Joseph B. Miles, Frank B. Miles, Stephen C. Milne, Joanna Mitro, Peter Monk, Susan Montgomery, John Moon, Marion Moore, John Morrison, B. Mortimer, Louise E. Moser, Wendell Motter, David E. Muller, Jan Mycielski, Victor Namias, David Newman, Morris Newman, Albert Nijenhuis, Zbigniew Nitecki, Ivan Niven, Karl K. Norton, Robert M. Norton, Don O'Shea, Barbara Osofsky, John Oxtoby, Edgard M. Palmer, Allan Parks, Alayne Parson, Emanuel Parzen, Gregg N. Patruno, C. Wayne Patty, Carl Pearson, N. Tenney Peck, Donald Pelletier, M. J. Pelling, Peter Perry, Richard E. Pfiefer, Richard S. Pierce, Walter V. Philipp, Vera Pless, Carl Pomerance, Horacio A. Porta, Stephen V. Portnoy, Geoff Price, J. R. Quine, Q. I. Rahman, Arlan Ramsey, Raymond Redheffer, Jim Reid, Edward M. Reingold, Clifford A. Reiter, Bruce Reznick, Gerhard Ringel, Joseph B. Roberts, Rosemary Roberts, Jack M. Robertson, Derek J. S. Robinson, Raphael M. Robinson, C. A. Rogers, Vijay Rohatgi, Kenneth Rosen, Alex Rosenberg, Joseph Rosenblatt, Kenneth A. Ross, Richard Roth, Joseph J. Rotman, Ranjan Roy, Halsey L. Royden, Walter Rudin, Hans Sagan, Yoram Sagher, Chih-Han Sah, Andre Scedrov, Margo Schaefer, Jonathan Schaer, Georg Schmidt, Glenn Schober, Lowell Schoenfeld, John Schue, Fred Schultz, Sy Schuster, Allen J. Schwenk, David B. Secrest, John L. Selfridge, George B. Seligman, Ernst S. Selmer, Jeffrey O. Shallit, Daniel Shanks, Harold N. Shapiro, Jay Shapiro, Louis Shapiro, Allen L. Shields, Mark Sheingorn, Carol Shilepsky, Don Shimamoto, Peter Shor, Yasutaka Sibuya, George Simmons, Gustavus Simmons, David Singer, David Singmaster, Neil J. A. Sloane, C. Small, Philip W. Smith, Craig Smorynski, Ernest Snapper, Alexia Sontag, Jerome Spanier, Saul Stahl, Ivar Stakgold, Dennis Stanton, Sherman Stein, John Steinig, Christine Stevens, B. M. Stewart, G. W. Stewart, Kenneth Stolarksy, John David Stone, Michael G. Stone, Richard G. Stoneham, Gilbert Strang, Lajos F. Takacs, Clifford Taubes, Alan D. Taylor, Carsten Thomassen, John Thorpe, G. E. Trapp, Jr., Hale Trotter, Alan Tucker, J. Jerry Uhl, Jr., Daniel Ullman, Stephen V. Ullom, Ron Umble, Ken Valente, M. Van Rossum-Wijsmuller, Kalathoor Varadarajan, Ilan Vardi, Richard Varga, Wolmer V. Vasconcelos, Rudolf Vyborny, Samuel S. Wagstaff, Jr., Ralph Walde, John H. Walter, Edward T. H. Wang, James Ward, Lawrence C. Washington, Michael Waterman, Frank Wattenberg, James R. Weaver, Paul M. Weichsel, Elliot C. Weinberg, Howard J. Weiner, Lutz Weis, Gregory P. Wene, John E. Wetzel, Dennis White, Sue Whitesides, W. Thurmon Whitley, Gareth Williams, Kenneth S. Williams, A. Rice Wolf, Frank Wolf, Derick Wood, Jay Wood, Michael Woodroofe, Robert E. Woodrow, Jang-Mei Gloria Wu, Adil Yaqub, Peter Yff, Elizabeth Yip, Lawrence Zalcman, Peter W. Zehna, Doron Zeilberger, Joseph L. Zemmer

TELEGRAPHIC REVIEWS

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T:	Textbook	P:	Professional Reading	1–4:	Semesters
C:	Computer Software	L:	Undergraduate Library	**:	Special Emphasis
S:	Supplementary Reading	13:	Grade Level	??:	Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

General, L. *What Game Is That?* Kenneth Watson-Jones. Vantage Pr, 1987, xiii + 58 pp, $8.95. [ISBN: 0-533-06966-1] For high school or college students. Forty-seven puzzles, problems, and games of the brain-teaser kind. Almost all will be known to aficionados but amateurs will find challenges. Nicely organized. Reviewer's copy was missing figure 40. Answer to number 12 is not complete. Statement of and answer to number 16 are wrong. Some nice philosophical touches in Foreward, Preface, and text itself. JK

Elementary, S*(13-14). *An Easy Course in Using the HP-28C.* John W. Loux, Chris Coffin. Grapevine, 1987, 233 pp, $22 (P). [ISBN: 0-931011-17-5] A very gentle introduction to the HP-28C calculator. The book is a programmed learner with problems interspersed within explanations. Topics covered include menues, editing, stack operations, postfix notation, and programming. This book would be ideal for a high school or undergraduate mathematics major who desires a working knowledge of HP-28C fundamentals. AM

Elementary, T(13). *Intermediate Algebra, Fourth Edition.* M.A. Munem, W. Tschirhart. Worth, 1988, xii + 612 pp, $28.95. [ISBN: 0-87901-377-X] Many excellent problems. Problem sets for each section/chapter are graded for difficult. Examples in text are clear (without assumptions) and easy to read. Page layout and type choices highlight important information. Includes discussions of systems of linear equations, sequences, series, and geometry. Highlights calculator usage. Good section on translating English expressions to mathematical expressions. (*First Edition*, TR, February 1973.) LB-E

Mathematics Appreciation, T(13). *Prerequisite Algebra.* Charles P. McKeague. Harcourt Brace Jovanovich, 1988, xv + 640 pp, $26. [ISBN: 0-15-571093-1] Author's stated goal is to make it possible in an intermediate algebra course to actually cover the latter topics (conic sections, logarithms, polynomial functions) by greatly condensing the review of elementary algebra. AWR

Precalculus, T(13: 1). *Precalculus.* Jerome E. Kaufmann. Prindle, Weber & Schmidt, 1988, xv + 569 pp, $29. [ISBN: 0-534-92007-1] Solid, complete-as-need-be, no-frills treatment of precalculus. Emphasis on problem solving, graphing techniques, and notion of function, all with an eye on the calculus. Plenty of worked-out examples, graded exercise sets, and study aids. Nicely laid out and printed, but not all figures ring true, especially parabolas and hyperbolas and cycloids (ordinary, hypo-, epi-). The curtate cycloid in Figure 8.24 is just plain wrong. Rightfully, the calculator is treated as the tool it is. JK

Precalculus, T(13). *Algebra with Trigonometry for College Students.* Charles P. McKeague. Harcourt Brace Jovanovich, 1988, xvii + 700 pp, $26. [ISBN: 0-15-502120-6] A well-written text covering algebra, conic sections, relations and functions, logarithms, and trigonometry. The chapters are organized into sections which can be presented in a 45-50 minute class lecture. Each section contains numerous exercises, including some application problems. In addition, each chapter ends with a chapter summary and a chapter test. Requires only a beginning algebra course as a prerequisite. RH

Precalculus, T(13). *Trigonometry.* John Baley, Martin Holstege. Random House, 1987, ix + 452 pp, $18. [ISBN: 0-394-35461-3] Definitely conversational in tone, including numerous inserts on using various calculators and little conversations by one who asks questions and one who answers them. Includes complex numbers and polar coordinates. AWR

Finite Mathematics, T(13). *Applied Mathematics for Business, Economics, and the Social Sciences, Third Edition.* Frank S. Budnick. McGraw-Hill, 1988, xxi + 1041 pp, $39.95. [ISBN: 0-07-008876-4] A text suitable for a two-term course combining finite mathematics with calculus. Intended for students of business, economics, and, to a lesser extent, social sciences. Prerequisite: high school algebra. Many

applications and examples, 3500 exercises and helpful algebra reminders at strategic locations throughout the text. Answers given for odd-numbered problems. (*First Edition,* TR, November 1979.) SM

Education, T(14-16). *Mathematics for Elementary Teachers: A Contemporary Approach.* Gary L. Musser, William F. Burger. Macmillan, 1988, xix + 874 pp. [ISBN: 0-02-384760-3] Development parallels typical sequence of topics in elementary school curriculum. Heavy emphasis on problem-solving strategies and many good problems, but claim to integrated problem-solving approach is belied by Course Options suggestion to de-emphasize or omit problem solving if time is short. Use of hands-on activities and bibliographies for each topic are confined to supplementary Student Resource Handbook. MW

Education, S*(15-16). *Middle Grades Mathematics Project.* Addison-Wesley, 1986, (P). *Mouse and Elephant: Measuring Growth.* Janet Shroyer, William Fitzgerald, 128 pp [ISBN: 0-201-21474-1]; *Factors and Multiples.* William Fitzgerald, *et al,* 152 pp [ISBN: 0-201-21475-X]; *Similarity and Equivalent Fractions.* Glenda Lappan, *et al,* 166 pp [ISBN: 0-201-21476-8]; *Spatial Visualization.* Mary Jean Winter, *et al,* 150 pp [ISBN: 0-201-21477-6]; *Probability.* Elizabeth Phillips, *et al,* 191 pp. [ISBN: 0-201-21478-4] Each book is a 2-3 mathematics unit for grades 5-8, emphasizing problem solving approach, use of concrete manipulatives, and carefully integrated and sequenced lessons. Detailed lesson plans include worksheets, transparency masters, and scripts for teacher talks and action. Three phases of instruction: launch challenge, class exploration, summarize. Both elementary and secondary methods students should see these. MW

Education, S(15-18). *Readings for Enrichment in Secondary School Mathematics.* Ed: Max A. Sobel. NCTM, 1988, v + 258 pp, $10 (P). [ISBN: 0-87353-252-X] The classics from two previous NCTM publications: eight articles from the Twenty-eighth Yearbook and five from *Topics for Mathematics Clubs.* Also eight reprinted from the *Mathematics Teacher* and three original articles. Many updated bibliographies. Suitable for independent study by talented students or as the focus of enrichment lessons. Excellent resource should be on the supplementary list for pre-service and in-service teachers. MW

Education. *Mirrors of Minds: Patterns of Experience in Educational Computing.* Ed: Roy D. Pea, Karen Sheingold. Ablex, 1987, xvii + 329 pp, $42.50; $19.95 (P). [ISBN: 0-89391-422-3] Collection of papers from Bank Street College Center for Children and Technology. Highlights learner and educational process, not machines, as starting points for studies of the role of computer technology in education. Examines the potential impact of computer-based technologies on educational innovation, the cognitive outcomes of learning LOGO programming, and the issues involved in designing software to promote critical inquiry. MW

Education, P. *Mathematics Tests Available in the United States and Canada.* James S. Braswell, Alicia A. Dodd. NCTM, 1988, v + 34 pp, $3.50 (P). [ISBN: 0-87353-253-8] A catalogue of standardized tests, arranged by grade level and subject, with information on publisher, purpose, published descriptions, and other relevant details. Includes directory of publishers and distributors, and a bibliography of reviews and descriptions. LAS

History, L*. *Izrail M. Gelfand: Collected Papers, Volume I.* Ed: S.G. Gindikin, *et al.* Springer-Verlag, 1987, vi + 883 pp, $149.50. [ISBN: 0-387-13619-3] First of three volumes of Gelfand's 450 research papers, from Banach algebras (normed rings) to cell biology. This volume contains survey papers by and about Gelfand; papers on Banach algebras, differential equations, and mathematical physics; a complete bibliography; and tentative contents of subsequent volumes. LAS

History, P*, L*. *The Probabilistic Revolution.* Ed: Lorenz Krüger, *et al.* MIT Pr, 1987, $60 set. *Volume 1: Ideas in History,* xv + 449 pp [ISBN: 0-262-11118-7]; *Volume 2: Ideas in the Sciences,* xvii + 459 pp. [ISBN: 0-262-11119-5] 35 essays arising from a year-long study undertaken by an international interdisciplinary research group convened to investigate the historical roots of the revolution that supplanted determinism with probabilistic models as the paradigm of physical and social science. *Volume 1* deals principally with the nineteenth century; *Volume 2* with applications to twentieth century science (psychology, economics, biology, physics). LAS

History, T(13-16), S, L. *The Historical Roots of Elementary Mathematics.* Lucas N.H. Bunt, Phillip S. Jones, Jack D. Bedient. Dover, 1988, xiii + 299 pp, $7.95 (P). [ISBN: 0-486-25563-8] Unabridged republication of a book published by Prentice-Hall in 1976 (TR, January 1977). Authors' corrections have been included in this edition. RH

History, L*. *The Collected Papers of Albert Einstein, Volume 1: The Early Years, 1879-1902.* Ed: John Stachel. Princeton U Pr, 1987, lxvi + 433 pp, $52.50 [ISBN: 0-691-08407-6]; *English Translation.* Transl: Anna Beck. Princeton U Pr, 1987, xxii + 196 pp, $22.50 (P). [ISBN: 0-691-08475-0] First of an anticipated thirty volumes constituting the definitive, comprehensive edition of all Einstein papers, letters, and related documents. This volume contains letters, school notebooks, exams, and related materials from Einstein's birth until his appointment as a patent clerk. Includes 50 recently discovered letters between Einstein and Mileva Marić, his fellow student at ETH and first wife, and a brief account of his childhood written in 1924 by his sister Maja Winteler-Einstein. Each manuscript is untranslated, typeset in a manner that is faithful to the original German; all editor's notes are in English. The accompanying paperback contains typewritten English translations of all original documents in the official edition. LAS

History, L. *Memoir of the Life and Labours of the*

Late Charles Babbage. H.W. Buxton. Ed: Anthony Hyman. Reprint Ser. for the History of Comput., V. 13. MIT Pr, 1988, xxi + 401 pp, $50. [ISBN: 0-262-02269-9] Thirteenth in a series on the history of computing. Buxton was a barrister (and self-taught in mathematics) whom Babbage entrusted to write his memoirs. The manuscript was not published in its time and remained in storage until recently. This edition provides a view of Babbage's broad range of interests and his own thoughts, as conveyed to Buxton, on his Analytical Engines. MR

History, P, L.** *Archimedes.* E.J. Dijksterhuis. Transl: C. Dikshoorn. Princeton U Pr, 1987, 457 pp, $15 (P); $50. [ISBN: 0-691-02400-6; 0-691-08421-1] A classic study now augmented with Wilbur R. Knorr's 23-page bibliographic essay "Archimedes after Dijksterhuis: A Guide to Recent Studies." Knorr's essay ends with a provocative postscript on "Who was Archimedes?" Knorr provides over 200 references to recent (several as late as 1987) studies on or related to Archimedes. The paperback edition is a real bargain. JK

History, T(15), S*, L*. *János Bolyai Appendix: The Theory of Space.* Ed: Ferenc Kárteszi. Math. Stud., V. 138. North-Holland (US Distr: Elsevier Science), 1987, 238 pp, $100 (P). [ISBN: 0-444-86528-4] Revision and translation of a memorial volume first published in 1952. Bolyai's *Appendix*—containing his ideas of non-Euclidean geometry—appears here in its original Latin as well as in translation. As the *Appendix* is extremely terse, editor Kárteszi has included a chapter of "explanatory, supplementary, orientative remarks." In addition to this, Kárteszi brackets the *Appendix* with chapters on the evolution of the concept of space and on further evolution of Bolyai's ideas. Beautifully written, this book illuminates the drama surrounding János Bolyai's life and work. LW

History, P*, L*.** *Pascal's Arithmetical Triangle.* A.W.F. Edwards. Oxford U Pr, 1987, xii + 174 pp, $37.50. [ISBN: 0-85264-283-0] A comprehensive history of Pascal's triangle, beginning with its distinct roots in recursion (the 'figurate' triangle), in combinations (the 'combinatoric' triangle), and in algebra (the 'binomial' triangle). Edwards, a reader in mathematical biology at Cambridge, identifies particular interpretations of early examples (Greek, Chinese, Hindu, etc.), and then shows how Pascal and his successors (Wallis, Newton, Bernoulli) linked diverse interpretations into a powerful tool. LAS

Logic, P. *Methods and Applications of Mathematical Logic.* Ed: Walter A. Carnielli, Luiz Paulo de Alcantara. Contemp. Math., V. 69. AMS, 1988, xi + 250 pp, $28 (P). [ISBN: 0-8218-5076-8] Proceedings of the Seventh Latin American Symposium on Mathematical Logic held July 29-August 2, 1985. Sixteen research papers on set theory, algebraic logic, philosophical aspects of logic, and interactions between logic, mathematics, and computer science. Includes summary of work of A.I. Arruda. KS

Foundations, T*(14: 1), S*, L. *Foundations of Higher Mathematics.* Peter Fletcher, C. Wayne Patty. Ser. in Math. Prindle, Weber & Schmidt, 1987, xiii + 263 pp, $28.50. [ISBN: 0-87150-164-3] Introduces basic techniques of writing proofs and fundamental ideas in the foundations of mathematics, logic, sets, induction, relations, functions, number systems, cardinality, and introductory algebra and analysis. Readable, with lots of worked examples, exercises, and illustrations. CEC

Graph Theory, P. *Recent Results in the Theory of Graph Spectra.* Dragos M. Cvetković, *et al.* Annals of Disc. Math., V. 36. North-Holland (US Distr: Elsevier Science), 1988, xi + 306 pp, $97.25. [ISBN: 0-444-70361-6] Results in the theory of graph spectra which appeared since 1978. Topics include graphs with least eigenvalue -2, distance regular graphs, graph polynomials, and spectra of infinite graphs. Most theorems stated with no proof, but over 700 references given. LC

Combinatorics, S(13-16), L.** *Challenging Mathematical Programs with Elementary Solutions.* A.M. Yaglom, I.M. Yaglom. Transl: James McCawley, Jr. Dover, 1987, $5.95 each (P). *Volume I: Combinatorial Analysis and Probability Theory,* ix + 231 pp [ISBN: 0-486-65536-9]; *Volume II: Problems From Various Branches of Mathematics,* ix + 214 pp. [ISBN: 0-486-65537-7] Republication of the translation of *Non-Elementary Problems in an Elementary Exposition* from the Russian. *Volume I* contains 100 problems with hints and detailed solutions dealing with combinatorial analysis (prerequisite: standard high school mathematics); problems concerning other areas are in *Volume II.* For problem-solving seminars, math clubs, etc., at the level of late high school/early college. RB

Combinatorics, P. *Combinatorics '86.* Ed: A. Barlotti, M. Marchi, G. Tallini. Annals of Disc. Math., V. 37. North-Holland (US Distr: Elsevier Science), 1988, xvi + 502 pp, $126.25 (P). [ISBN: 0-444-70369-1] This volume contains the proceedings of the conference held at the Centro di cultura of the Università Cattolica of Milano, in Passo della Mendola from June 30 to July 5, 1986. The focus is on combinatorial and incidence geometry and its links with geometry, graph theory, algebraic structures, and applications to coding theory and computer science. CEC

Discrete Mathematics, T(14-15: 1, 2), L. *Discrete Mathematics.* Kenneth P. Bogart. DC Heath, 1988, xvii + 834 pp, $29. [ISBN: 0-669-08665-7] More demanding of the student and more comprehensive than most of the new books in discrete mathematics, there is ample material for a year's course. In addition to more details on the normal topics, attention is paid to logical quantifiers, growth rates, matrix algebra, probability, and abstract algebra, including coding theory. Good selection of exercises with answers and solutions, suggested readings, appendix on generating functions. JS

Discrete Mathematics, T(13-14: 1, 2). *Introduction to Discrete Mathematics.* James Bradley.

Addison-Wesley, 1988, xiv + 685 pp, $40.95. [ISBN: 0-201-10628-0] Covers standard topics with algorithms as the unifying theme. Over 40 algorithms in Pascal pseudo-code are presented, with complexity discussed in many of them. Other topics include elementary number theory, probability models, and nearly 50 pages on elementary matrix theory (Gaussian elimination, inverses). LC

Number Theory, P. *Lecture Notes in Mathematics-1290: Diophantine Approximation and Transcendence Theory.* Ed: G. Wüstholz. Springer-Verlag, 1987, 243 pp, $21.20 (P). [ISBN: 0-387-18597-6] Formulated in part during a 1985 workshop at Bonn, this volume contains articles on the Hasse principle for cubic surfaces, heights on Abelian varieties, large transcendence degree, linear forms in logarithms, and the Thue-Mahler equation. GG

Number Theory, T(18: 1), S, P. *Lectures on a Method in the Theory of Exponential Sums.* M. Jutila. Springer-Verlag, 1987, viii + 131 pp, $12 (P). [ISBN: 0-387-18366-3] Self-contained treatment of exponential sums arising in the study of the Riemann zeta function and, more generally, of Dirichlet series arising from modular forms. The first two of the book's four chapters cover summation formulas and exponential integrals. These are applied in the final two chapters to obtain general and specific transformation formulas, including many new results or new proofs. GG

Number Theory, T(18: 1), S, P. *Hilbert Modular Surfaces.* Gerard van der Geer. Ergebnisse der Math. und ihrer Grenzgebiete, B. 16. Springer-Verlag, 1988, ix + 291 pp, $69.50. [ISBN: 0-387-17601-2] To an ever-increasing extent, solutions to Diophantine problems rely on the power of algebraic geometry. Many times there is an explicit connection with automorphic forms. There are also conjectures relating automorphic forms to prime ideal decompositions in number fields. A Hilbert modular surface is a quotient of the product of two upper half planes by a discrete group of two-by-two matrices over a real quadratic field. The book studies the geometry of these surfaces and the automorphic forms on them. GG

Number Theory, S(15-17), P, L.** *Elementary Theory of Numbers.* W. Sierpiński. Math. Lib., V. 31. North-Holland (US Distr: Elsevier Science), 1988, xii + 513 pp, $105.25. [ISBN: 0-444-86662-0] Few changes, other than references, to the 1964 original in this second edition, edited by A. Schinzel. Similar to *The Theory of Numbers* by Hardy and Wright, this book considers primarily Diophantine and divisibility problems whose solutions lie within the ring of integers. A small sample of the range: Pell's equation, distribution of primes, congruences, Mersenne and Fermat numbers, and Waring's problem. GG

Number Theory, L.** *The Book of Prime Number Records.* Paulo Ribenboim. Springer-Verlag, 1988, xiii + 476 pp, $49.80. [ISBN: 0-387-96573-4] An interesting and truly delightful treatment of prime numbers. In spite of the title, this is much more than a book of records. It contains up-to-date historical presentations of many important problems concerning prime numbers, including some proofs. Several topics are covered, for example the distribution of primes and primality testing. Of course, the relevant records are also included. Written in a very readable manner which should be accessible and interesting to both specialists and nonspecialists. RH

Number Theory, T*(14-16: 1), S, L*. *Elementary Number Theory and Its Applications, Second Edition.* Kenneth H. Rosen. Addison-Wesley, 1988, xiii + 466 pp, $37.95. [ISBN: 0-201-11958-7] New to this edition (*First Edition,* TR, January 1986) are sections on basic axioms for integers and mathematical induction, more routine exercises, simplified proofs, earlier introduction of Fibonacci numbers, expanded coverage of big-0 notation, sections on sum of squares and on factoring using continued fractions, recent discoveries of large primes, and biographies of mathematicians. RH

Number Theory, T(17-18: 1), S, P, L. *An Introduction to the Theory of the Riemann Zeta-Function.* S.J. Patterson. Stud. in Adv. Math., V. 14. Cambridge U Pr, 1988, xiii + 156 pp, $34.50. [ISBN: 0-521-33535-3] The Riemann zeta-function is a meromorphic function which can be studied through complex analysis, but on the other hand yields interesting results concerning prime numbers. This book presents a classical introduction to the Riemann zeta-function, including an historical introduction and a lengthy discussion of the Riemann Hypothesis. Contains many exercises and includes appendices which summarize some of the background mathematics needed. RH

Linear Algebra, T*(16-17: 1), L. *A Second Course in Linear Algebra.* William C. Brown. Wiley, 1988, x + 264 pp, $34.95. [ISBN: 0-471-62602-3] Preparation for graduate work in algebra, analysis, topology, applied mathematics. Presumes the usual one-semester undergraduate course. Studies familiar topics more deeply, then proceeds to multilinear algebra, canonical forms of matrices, normed linear spaces, inner-product spaces. Deliberately formal; excellent problem sets. DFA

Linear Algebra, T(14: 1). *Linear Algebra with Applications.* Charles G. Cullen. Scott Foresman, 1988, 393 pp, $26.50. [ISBN: 0-673-18570-2] Covers standard topics with emphasis on computational aspects. Applications include linear programming, graph theory, least square approximations, and quadratic forms. LC

Linear Algebra, T*(14: 1). *A Plaid Linear Algebra.* Robert Messer (Dept. of Math., Albion Coll., Albion, MI 49224), 1988, viii + 189 pp, $22 (P). An interesting textbook emphasizing general theory over computations. Covers the basics together with some applications. Specifically designed to help students develop the mathematical maturity needed for subsequent mathematics courses. AO

Linear Algebra, T(14-15: 1). *Linear Algebra.* Carroll Wilde. Addison-Wesley, 1987, xvii + 492

pp, $39.75. [ISBN: 0-201-13089-0] Quite similar in content and organization to other recent linear algebra texts; begins with linear systems, matrix algebra, and determinants before considering vector spaces and transformations. Several distinguishing features include good graphics, review checklists, and use of the computer. Applications are treated in some detail. JS

Linear Algebra, T(14: 1). *Elementary Linear Algebra.* Roland E. Larson, Bruce H. Edwards. DC Heath, 1988, xiii + 533 pp, $29. [ISBN: 0-669-14583-1] Chapters on systems of linear equations, matrices, and determinants precede coverage of vector spaces and linear transformations. Applications in areas such as computer graphics, cryptography, differential equations, etc., are included in each chapter. A final chapter covers numerical methods. A diskette of software operating on an IBM-PC is available. JNC

Group Theory, P. *Whitehead Groups of Finite Groups.* Robert Oliver. Math. Soc. Lect. Note Ser., V. 132. Cambridge U Pr, 1988, 349 pp, $34.50 (P). [ISBN: 0-521-33646-5] The intent of the author is to draw together a variety of results scattered through the literature and dealing with recent work on Whitehead groups. Part I is general theory, Part II treats group rings on p-groups, and Part III considers general finite groups. Bibliography, index. JS

Group Theory, P. *The Arcata Conference on Representations of Finite Groups.* Ed: Paul Fong. Proc. of Symposia in Pure Math., V. 47. AMS, 1987, $122 set [ISBN: 0-8218-1479-6]. *Part 1,* xi + 487 pp; *Part 2,* viii + 552 pp. A collection of 29 expository papers and 69 research papers presented at the Summer Research Institute held at Humboldt State University in Arcata, California, July 7-25, 1986. Covers advances made in the representation theory of finite groups since the 1979 Santa Cruz Summer Institute on finite groups. RH

Algebra, P. *Commutative Algebra and Combinatorics.* Ed: M. Nagata, H. Matsumura. Adv. Stud. in Pure Math., V. 11. North-Holland (US Distr: Elsevier Science), 1987, 359 pp, $163.25. [ISBN: 0-444-70314-4] A collection of 24 papers in those aspects of commutative algebra which are particularly useful in combinatorics. The papers cover such topics as Cohen-Macaulay rings, representations of classical groups, and non-commutative invariant theory. SG

Algebra, T(17-18: 1), P. *Model Theory and Modules.* Mike Prest. Math. Soc. Lect. Note Ser., V. 130. Cambridge U Pr, 1988, xviii + 380 pp, $37.50 (P). [ISBN: 0-521-34833-1] A monograph on the interaction between the two mathematical fields, concerned with applications of model theory to the study of modules (e.g., relationship between pp-types and pure-injective modules) and with use of modules as examples in stability theory. The book is intended to be accessible by graduate students and established workers in algebra or logic. RB

Algebra, P. *Triangulated Categories in the Representation Theory of Finite Dimensional Algebras.* Dieter Happel. Math. Soc. Lect. Note Ser., V. 119. Cambridge U Pr, 1988, ix + 208 pp, $24.95 (P). [ISBN: 0-521-33922-7] Describes how triangulated categories can be used in studying modules over finite dimensional K-algebras, where K is a field which is usually assumed to be closed. Also shows that when recent methods of representation theory are applied, certain triangulated categories become more accessible. RH

Algebra, T(18: 2). *Real Reductive Groups I.* Nolan R. Wallach. Pure & Appl. Math., V. 132. Academic Pr, 1988, xix + 412 pp, $59.95. [ISBN: 0-12-732960-9] The representation theory of real reductive groups is almost completely the product of Harish-Chandra. The author attempts to present the algebraic aspects of the theory in this volume and the analytic aspects in the next. The heart of this volume is the theory of the real Jacquet module and the asymptotic behavior of matrix coefficients. Suitable for an advanced graduate course. No exercises. MR

Algebra, T(16-17: 2), P. *Modular Lie Algebras and Their Representations.* Helmut Strade, Rolf Farnsteiner. Pure & Appl. Math., V. 116. Marcel Dekker, 1988, viii + 301 pp, $79.75. [ISBN: 0-8247-7594-5] Modular Lie algebras are Lie algebras over fields of positive characteristics. Although the last several decades have seen "vigorous development" in the theory of modular Lie algebras, these algebras have yet to be completely classified. This book first sketches the classical theory of Lie algebras over fields of characteristic 0, and then proceeds to introduce the structure and representation theory of modular Lie algebras. The only prerequisite for the first part of the book is a knowledge of linear algebra; for the chapters addressing representation theory, the reader needs some familiarity with the techniques from the theory of associative algebras. Although written quite formally, the book includes exercises and could be used as the text for a course. LW

Calculus. *Calculus with Analytic Geometry, Fourth Edition.* Earl W. Swokowski. Prindle, Weber & Schmidt, 1988, xii + 914 pp, $39.50. [ISBN: 0-87150-007-8] Intended for a two-term freshman calculus course. Covers standard topics. Many exercises, both standard plug-n-chug and story problems. Applications from a wide variety of fields. Answers to odd numbered problems. Well illustrated. All major theorems proved. (*First Edition,* TR, December 1975; *Second Edition,* TR, June-July 1979; *Third Edition,* TR, November 1984.) SM

Calculus, T(13: 1). *Brief Calculus for Management and the Life and Social Sciences.* Donald L. Stancl, Mildred L. Stancl. Richard D Irwin, 1988, xv + 505 pp, $33.95. [ISBN: 0-256-06303-6] Use of functions to construct models, and calculus to analyze models in management, economics, and sciences. Heavy emphasis on drawing and interpreting graphs of functions. Derivation, integration, exponential and logarithmic functions, and functions of several variables. Separate computer manual with IBM PC disk in Basic illustrating the material is available. MS

Calculus, T. *Calculus for Management and the Life*

and Social Sciences. Donald L. Stancl, Mildred L. Stancl. Richard D Irwin, 1988, xv + 684 pp, $35.95. [ISBN: 0-256-05714-1] Presents basic concepts and techniques of calculus with an emphasis on the use of functions to construct models in management, economics, and the sciences. A companion computer manual containing an IBM PC formatted disk with programs illustrating text material is available. JNC

Real Analysis, T*(16-17: 2), P, L*. *Real Variables.* Alberto Torchinsky. Addison-Wesley, 1988, xii + 403 pp, $43.25. [ISBN: 0-201-15675-x] Covers standard topics on Lebesque measure and integration (in R^n) as well as some of the basics of functional analysis. Assumes "a good advanced calculus background as well as an elementary knowledge of the theory of metric spaces." Very well motivated with extensive set of problems at end of each chapter. BH

Complex Analysis, P. *Lecture Notes in Mathematics-1295: Séminaire d'Analyse, P. Lelong—P. Dolbeault—H. Skoda.* Ed: P. Lelong, P. Dolbeault, H. Skoda. Springer-Verlag, 1987, vii + 283 pp, $25.80 (P). [ISBN: 0-387-18691-3] Collection of 11 articles in several complex variables (in French). Sequel to *Lecture Notes in Mathematics 1028, 1198.* BH

Complex Analysis, P. *Branched Coverings and Algebraic Functions.* Makoto Namba. Pitman Res. Notes in Math. Ser., V. 161. Longman Scientific & Technical (US Distr: Wiley), 1987, 201 pp, $49.95 (P). [ISBN: 0-470-20949-6] Covers theory of finite branched coverings of projective complex manifolds in connection with the theory of algebraic functions of several complex variables. Two main problems are discussed. First, give conditions for the existence of a Galois (Abelian) covering which branches at a divisor D, and second, describe the set of all finite Galois (Abelian) covering which branch at most at D. The first of these is Fenchel's problem, and the second is class field theoretic and can be traced back to Hilbert and Weil. MR

Differential Equations, P. *Fuchsian Differential Equations: With Special Emphasis on the Gauss-Schwarz Theory.* Masaaki Yoshida. Aspects of Math., E11. Freidr Vieweg & Sohn, 1987, xiv + 215 pp, (P). [ISBN: 3-528-08971-7] A study in linear ordinary differential equations and systems of partial differential equations with finite dimensional solution spaces. Concentrates on Gauss-Schwarz theory of the hypergeometric differential equation, in particular as part of the theory of orbifold. The main question is how to constuct a differential equation (in several variables) with a given set of singularities. A restricted answer is given. MR

Differential Equations, P. *Oscillation Theory of Differential Equations with Deviating Arguments.* G.S. Ladde, V. Lakshmikantham, B.G. Zhang. Pure & Appl. Math., V. 110. Marcel Dekker, 1987, vi + 308 pp, $89.75. [ISBN: 0-8247-7738-7] An ordinary differential equation with deviating arguments (ODEWDA) is one in which the solution depends on the past history and not just on the current state; [e.g., $2y''(t) + y(t) - y(t-\pi) = 0, t \geq 0$, has $y(t) = 1 - \sin t$ as a solution]. There are six chapters: preliminaries, first order linear, first order nonlinear, second order, higher order, and systems of differential equations; over 300 references are listed. It is "the first English language book that offers a systematic study of the theory of oscillation of ODEWDA." RSF

Partial Differential Equations, P. *First International Symposium on Domain Decomposition Methods for Partial Differential Equations.* Ed: Roland Glowinski, *et al.* SIAM, 1988, x + 431 pp. [ISBN: 0-89871-220-3] Proceedings of a symposium in Paris, January 1987. Includes 22 papers, and the list of contributors. Topics include elliptic problems in two- and three-dimensions, mixing finite elements and finite differences, preconditioning, fluid dynamics, structural analysis, incompressible fluid flows, vortex subdomains, and transonic flows. RSF

Partial Differential Equations, T(16-17: 1), L. *Numerical Solution of Partial Differential Equations by the Finite Element Method.* Claes Johnson. Cambridge U Pr, 1987, 278 pp, $69.50; $24.95 (P). [ISBN: 0-521-345-146; 0-521-347-580] An introduction to the finite element method. Covers elliptic, parabolic, and hyperbolic equations. Also includes a chapter on finite element methods for integral equations and an introduction to some nonlinear problems. The material concerning parabolic and hyperbolic problems contains some recent developments which have not appeared in earlier texts. Includes appropriate exercises. Prerequisites are advanced calculus and linear algebra. RH

Partial Differential Equations, P. *Lecture Notes in Mathematics-1297: Numerical Methods for Partial Differential Equations.* Ed: Zhu You-lan, Guo Ben-yu. Springer-Verlag, 1987, xi + 244 pp, $25.80 (P). [ISBN: 0-387-18730-8] Proceedings of a conference held in Shanghai, March 1987. Includes sixteen papers (in English) of the 75 papers presented (all 75 are listed by author). Topics include Hamiltonian equations, Signorini problems, nonlinear wave equations on n-dimensional spheres, parallel algorithms and domain decomposition, degenerate second order hyperbolic equations, error expansions for finite element approximations, the MKDV equation, nonlinear fourth order equations, parabolic systems, discontinuous vortical flows, folds of degree four and swallowtail catastrophe, viscous splitting in bounded domains, inverse scattering, and quasilinear hyperbolic initial-boundary-value problems. RSF

Partial Differential Equations, P. *Sensors and Controls in the Analysis of Distributed Systems.* A. El Jai, A.J. Pritchard. Math. & Its Applic. Halsted Pr, 1988, 125 pp, $64.95. [ISBN: 0-470-21023-0] A theoretical study of the structure of input and output operators (i.e., controls and sensors) in infinite-dimensional distributed parameter systems (example application: heat flow in a furnace). The study, rigorously approached via semigroups, is limited to systems described by parabolic partial differential equations; a final chapter considers hyperbolic and non-

linear systems. RB

Numerical Analysis, P. *Advanced Boundary Element Methods.* Ed: T.A. Cruse. Intern. Union of Theoret. & Appl. Mech. Springer-Verlag, 1988, xix + 489 pp, $83.70. [ISBN: 0-387-17454-0] Fifty-two papers from a symposium on boundary (as opposed to finite) element methods. Diverse mechanical applications are considered. BC

Numerical Analysis, P. *Numerical Methods.* Ed: D. Greenspan, P. Rózsa. North-Holland (US Distr: Elsevier Science), 1988, 688 pp, $223.75. [ISBN: 0-444-70209-1] Proceedings of the Conference on Numerical Methods, organized by the János Bolyai Mathematical Society, held at the campus of the Technical University for Heavy Industry in Miskolc, Hungary in 1986. Includes 5 invited papers and 36 refereed papers from the conference arranged into the following sections: linear algebra, differential equations, applications, nonlinear equations and optimization, further topics in numerical mathematics, and program packages. RH

Functional Analysis, P. *Aspects of Positivity in Functional Analysis.* R. Nagel, U. Schlotterbeck, M.P.H. Wolff. Math. Stud., V. 122. North-Holland (US Distr: Elsevier Science), 1986, xiii + 277 pp, $40.75 (P). [ISBN: 0-444-87959-5] Proceedings of a conference in honor of H.H. Schaefer held at Tübingen in 1985. The general theme is positivity in functional analysis, with eleven invited addresses and thirteen shorter papers. List of participants. JS

Functional Analysis, T(15: 1). *Metric Spaces.* E.T. Copson. Tracts in Math., No. 57. Cambridge U Pr, 1988, vii + 143 pp, $12.95 (P). [ISBN: 0-521-35732-2] Leisurely treatment of metric spaces. Applications of theory to classical algebra and analysis. Knowlege of elements of theory of uniform convergence is assumed. Includes open and closed sets, complete metric spaces, connected sets, compactness, functions and mappings, and some applications. Banach and Hilbert spaces introduced at the end along with general topology. Corrected reprint of 1968 original (TR, November 1969). MS

Functional Analysis, P. *Perturbation Bounds for Matrix Eigenvalues.* Rajendra Bhatia. Res. Notes in Math. Ser., V. 162. Longman Scientific & Technical (US Distr: Wiley), 1987, 129 pp, $41.95 (P). [ISBN: 0-470-20917-8] The central question is: If A and B are $n \times n$ matrices, how close are the eigenvalues of A and B given the distance $\|A - B\|$? This monograph establishes good bounds on the eigenvalue distance for various classes of matrices. MR

Analysis, T(17-18: 1, 2), S*, P, L**.** *Fourier Analysis.* T.W. Kōrner. Cambridge U Pr, 1988, xii + 591 pp, $95. [ISBN: 0-521-25120-6] An exceptionally well-written and wide-ranging account of the theory and applications of Fourier analysis. Interspersed with down-to-earth examples and anecdotes (such as the author's description of being scalded by the time delay equation of a dormitory shower), this book is rewarding even just to browse through. It is one of the few $95 books that is actually worth the price. BC

Analysis, S(18), P, L.** *Lectures on Bochner-Riesz Means.* Katherine Michelle Davis, Yang-Chun Chang. London Math. Soc. Lect. Note Ser., V. 114. Cambridge U Pr, 1987, ix + 150 pp, $22.95 (P). [ISBN: 0-521-31277-9] "Self-contained exposition of the geometric theory of Bochner-Riesz means." Assumes graduate course in real analysis and basic familiarity with distributions and the Fourier transform. Wonderfully motivated with many intuitive insights. A joy to read. BH

Analysis, P. *Some Applications of Weighted Sobolev Spaces.* Alois Kufner, Anna-Margarete Sändig. Teubner-Texte zur Math., B. 100. BG Teubner Leipzig, 1987, 268 pp, 28 M (P). [ISBN: 3-322-00426-0] Sequel to first author's book *Weighted Sobolev Spaces* (Volume 31 in series). Part I studies elliptic boundary value problems for domains with conical corners and with edges. Part II deals with existence theorems for elliptic differential equations and problems of the type of degenerate equations and equations with singular coefficients. BH

Analysis, P. *Proceedings of the Analysis Conference, Singapore 1986.* Ed: Stephen T.L. Choy, Judith P. Jesudason, P.Y. Lee. Math. Stud., V. 150. North-Holland (US Distr: Elsevier Science), 1988, xii + 304 pp, Dfl. 175 (P). [ISBN: 0-444-70341-1] Collection of seventeen papers primarily in harmonic and functional analysis. BH

Analysis, P. *Moduli of Smoothness.* Z. Ditzian, V. Totik. Ser. in Computat. Math., V. 9. Springer-Verlag, 1987, ix + 227 pp, $54.90. [ISBN: 0-387-96536-X] The classical modulus of smoothness in approximation theory is defined by $\omega^r(f,t) = \sup_{0<h\leq t} \| \Delta_n^r f \|$. The authors introduce a new modulus $\omega_\varphi^r(f,t)_p = \sup_{0<h\leq t} \| \Delta_n^r \varphi f \|_L^p$ and apply it to problems in polynomial approximation and characterization of k-functionals. Results presented are new. Very clear presentation. BH

Analysis, P. *Nonlinear Analysis and Applications.* Ed: V. Lakshmikantham. Lect. Notes in Pure & Appl. Math., V. 109. Marcel Dekker, 1987, xix + 649 pp, $99.75 (P). [ISBN: 0-8247-7810-3] A collection of 84 papers presented at the Seventh International Conference on Nonlinear Analysis and Applications held at the University of Texas at Arlington in 1986. AWR

Algebraic Geometry, S(18), P. *Convex Bodies and Algebraic Geometry: An Introduction to the Theory of Toric Varieties.* Tadao Oda. Ergebnisse der Math. und ihrer Grenzgebiete, B. 15. Springer-Verlag, 1988, viii + 212 pp, $78. [ISBN: 0-387-17600-4] A geometric introduction to complex analytic spaces, concentrating on toric varieties, which are also called torus embeddings. Although the full strength of algebraic geometry is not required, knowledge of basic properties of algebraic varieties over the complex numbers is necessary. Many results, including those of the author, are less than fifteen years old. Discouraging price. GG

Algebraic Geometry, P. *Lecture Notes in Mathe-*

matics-1293: The Monodromy Groups of Isolated Singularities of Complete Intersections. Wolfgang Ebeling. Springer-Verlag, 1987, xiv + 153 pp, $17.30 (P). [ISBN: 0-387-18686-7] Arithmetical characterization (in terms of the Milnor lattice) for monodromy groups and vanishing cycles of certain intersection singularities. A step beyond hypersurfaces. BC

Algebraic Geometry, P. *Etale Cohomology and the Weil Conjecture.* Eberhard Freitag, Reinhardt Kiehl. Ergebnisse der Math. und ihrer Grenzgebiete, B. 13. Springer-Verlag, 1988, xviii + 317 pp, $98. [ISBN: 0-387-12175-7] A self-contained treatment of the etale cohomology necessary for Deligne's proof of the Weil conjectures. Including a chapter on the monodromy theory of Lefschetz pencils, a description of some applications of the conjectures, and a history of the Weil conjectures written by Dieudonné. SG

Algebraic Geometry, P. *Topics on Real and Complex Singularities: An Introduction.* Alexandru Dimca. Adv. Lect. in Math. Friedr Vieweg & Sohn, 1987, xvii + 242 pp, (P). [ISBN: 3-528-08999-7] An introduction to local singularity theory with special attention paid to the complex analytic situation and connections with algebraic geometry. Main topic covered is the finite determinancy of germs (alias local C-algebras). Presupposes basics of smooth manifolds, commutative algebra, and algebraic geometry. Takes the reader from introductory material through unpublished research work. Some exercises, examples of open problems, bibliography, and index. MR

Differential Geometry, P. *Foliations on Riemannian Manifolds.* Philippe Tondeur. Universitext. Springer-Verlag, 1988, xi + 247 pp, $29.80 (P). [ISBN: 0-387-96707-9] Foliations are generalizations of integral flows of dynamical systems to higher dimensions. This text represents notes from lectures given at the University of Illinois. Chapters 1-4 are an introduction to the subject of foliations, while the remainder of the book covers the relationship between foliations on Riemannian manifolds. No exercises. MR

Differential Geometry, T(17-18: 1), L.** *Differential Geometry, Gauge Theories, and Gravity.* M. Göckeler, T. Schücker. Mono. on Math. Physics. Cambridge U Pr, 1987, xii + 230 pp, $49.50. [ISBN: 0-521-32960-4] Introduction to those areas of modern differential geometry fundamental to particle physics and general relativity. Topics include Yang-Mills theories, gravity, fiber bundles, monopoles, instantons, spinors, and anomalies. AM

Geometry, P, L**.** *Shaping Space: A Polyhedral Approach.* Ed: Marjorie Senechal, George Fleck. Birkhauser Boston, 1988, xx + 284 pp, $49.95. [ISBN: 0-8176-3351-0] A polyhedral anthology in the Design Science Collection series edited by Arthur L. Loeb. Inspired by the Shaping Space Conference at Smith College in April 1984. Theory and history of polyhedra. Applications in science and engineering. For "all scholars and educators interested in and working with two- and three-dimensional structures and patterns." Richly illustrated (photographs, line and half-tone drawings) interdisciplinary guide to three-dimensional forms focusing on polyhedra. Chapters on model building, unsolved problems, polyhedra in the curriculum. An abundance of references in art, architecture, science and, even, geometry which will be unfamiliar to many mathematicians. Recommended for every school, college, and university library. JK

Algebraic Topology, P. *Lecture Notes in Mathematics-1298: Algebraic Topology, Barcelona 1986.* Ed: J. Aguadé, R. Kane. Springer-Verlag, 1987, x + 255 pp, $25.80 (P). [ISBN: 0-387-18729-4] Collected refereed papers from the Second Barcelona Conference on Algebraic Topology held in April 1986, and subsequently dedicated to the memory of Alex Zabrodsky. Consists of 15 main talks and 23 short communications on a variety of topics. LW

Differential Topology, P. *A Fête of Topology: Papers Dedicated to Itiro Tamura.* Ed: Y. Matsumoto, T. Mizutani, S. Morita. Academic Pr, 1988, xii + 602 pp, $60. [ISBN: 0-12-480440-3] A collection of terse and formal papers contributed by 30 authors. The papers are arranged into five categories: foliations, characteristic classes, singularities and orbifolds, 3- and 4-dimensional manifolds, and group actions. LW

Differential Topology, P. *Topics in Differential Geometry, V. I-II.* Ed: J. Szenthe, L. Tamássy. North-Holland (US Distr: Elsevier Science), 1988, $315.75 set [ISBN: 0-444-70090-0]. *Volume I,* 708 pp; *Volume II,* 666 pp. Proceedings of a conference held at Hajduszoboszló in August 1984. A variety of topics covered with emphasis on Finsler geometry, Riemannian geometry, and applications. MR

Differential Topology, S(16-17), P, L. *Models of the Real Projective Plane: Computer Graphics of Steiner and Boy Surfaces.* François Apéry. Comput. Graphics & Math. Models. Friedr Vieweg & Sohn, 1987, xi + 156 pp, (P). [ISBN: 3-528-08955-5] Nicely illustrated text (with 64 color plates) on immersions of P^2 in R^3. Hilbert thought you couldn't do it without introducing singularities. Boy, was he wrong. BC

Differential Topology, P. *Stratified Morse Theory.* Mark Goresky, Robert MacPherson. Ergebnisse der Math. und ihrer Grenzgebiete, B. 14. Springer-Verlag, 1988, xv + 272 pp, $75. [ISBN: 0-387-17300-5] Morse theory is generalized to stratified spaces, i.e., spaces with a decomposition into submanifolds (of various dimensions) so that singularities are locally constant along each submanifold. Applications are given, primarily to the topology of complex analytic varieties. The proofs of the main theorems, announced in 1980, appear here for the first time after a period of simplification and geometrization. RB

Topology, T(18), P, L. *Extensions and Absolutes of Hausdorff Spaces.* Jack R. Porter, R. Grant Woods. Springer-Verlag, 1987, xiii + 856 pp, $89. [ISBN: 0-387-96212-3] An extension of a Hausdorff space X is a space in which X is dense; the Iliadis absolute of X is an extremally disconnected zero-dimensional Hausdorff space EX that admits a per-

fect irreducible θ-continuous surjection onto X. Although conceptually quite different, extensions and absolutes can be constructed using similar tools. This tome is primarily a systematic study of extensions and absolutes of a Hausdorff space. Each chapter begins with a well-written and compelling introductory section, presenting the motivation and goals of the chapter. The text is nicely organized and pleasant to read. Each chapter is followed by numerous exercises, so the book would be ideal as a text for a graduate-level course. LW

Operations Research, P. *Traffic Processes in Queueing Networks: A Markov Renewal Approach.* Ralph L. Disney, Peter C. Kiessler. Math. Sci., V. 4. Johns Hopkins U Pr, 1987, xix + 251 pp, $49.50. [ISBN: 0-8018-3454-6] Study of queueing traffic process problems, including dependencies between traffic processes, e.g., cross-correlational properties. Arrival processes, overflow processes and finite capacity queues, Markov renewal processes, point processes, reversibility and traffic sets, detailed study of processes in small networks. RM

Operations Research, T(15-17: 1, 2), S. *Principles of Operations Research for Management, Second Edition.* Frank S. Budnick, Dennis McLeavey, Richard Mojena. Richard D Irwin, 1988, xx + 988 pp, $42.95. [ISBN: 0-256-02643-2] Textbook for a course in management science, operations research, or quantitative methods offered at schools of management/business administration, colleges, or universities. In this edition: coverage of mathematical models is unlinked from coverage of their solution algorithms for greater flexibility; extended end-of-chapter case studies added; answers provided for selected exercises. RB

Optimization, T(15-16: 1), L*. *The Mathematics of Nonlinear Programming.* Anthony L. Peressini, Francis E. Sullivan, J.J. Uhl, Jr. Undergrad. Texts in Math. Springer-Verlag, 1988, x + 273 pp, $42. [ISBN: 0-387-96614-5] After a brief review of unconstrained optimization, the text begins with the study of convex sets. Iterative and least squares methods are then treated in more detail than usual in the prerequisite analysis and linear algebra courses. The heart of the book uses geometry, penalty functions, Lagrange multipliers, and the simplex method in constrained optimization problems. A recurring theme is the Karush-Kuhn-Tucker existence theorem for convex programming. Each chapter concludes with a variety of exercises. Well-written, but several misprints. GG

Optimization, T(17: 2), P. *Maximum and Minimum Principles: A Unified Approach, with Applications.* M.J. Sewell. Cambridge U Pr, 1987, xvi + 468 pp, $79.50; $34.50 (P). [ISBN: 0-521-33244-3; 0-521-34876-5] Subject matter is optimization focused on nonlinear programming and on calculus of variations, but a nonstandard approach seeks unifying principles in the context of a saddle functional $L[x,u]$ concave in x, convex in u, with x in one inner product space, u in another. Chapter 1: Saddle function problems; Chapter 2: Duality and Legendre transformations; and Chapter 3: Upper and lower bounds via saddle functions. Said to be suitable for a final undergraduate year course, but students might find it stiff going. Last two chapters intended as reference for workers in the field. Exceptionally detailed table of contents enhances value as a reference book. AWR

Optimization, T(17), P. *Stability and Perfection of Nash Equilibria.* Eric van Damme. Springer-Verlag, 1987, xvii + 318 pp, $99. [ISBN: 0-387-17101-0] Solutions to game theory problems usually assume rational behavior on the part of one's opponent. Irrational behavior (non-cooperative game theory) is commonly covered in discussions of Nash equilibrium, a concept increasingly found to have certain limitations. In this book, an elaboration of the author's "refinements of the Nash equilibrium concept" that includes chapters on specific examples, various refinements of Nash equilibrium, are discussed. AWR

Dynamical Systems, S(17-18), P. *Dynamical Systems Approaches to Nonlinear Problems in Systems and Circuits.* Ed: Fathi M.A. Salam, Mark L. Levi. SIAM, 1988, xvi + 413 pp. [ISBN: 0-89871-218-1] Abstracts and papers contributed to illuminate areas of applied mathematics and engineering for which dynamical systems are appropriate. Specific areas include nonlinear circuits and systems, control systems, solids and vibrations (especially chaotic motion), and mechanical systems (e.g., robot manipulators). Two chapters of background papers on the general theory introduce the book. LB-E

Probability, S(16-18), P, L. *Counterexamples in Probability.* Jordan M. Stoyanov. Prob. & Math. Stat. Wiley, 1988, xxiii + 313 pp, $81.95. [ISBN: 0-471-91649-8] First two-thirds presents counterexamples in probability theory, including sections on random events, random variables, distributions, expectations, and limit theorems. Last third gives counterexamples in stochastic processes. Note price! RSK

Probability, P. *Ten Lectures on the Probabilistic Method.* Joel Spencer. CBMS-NSF Reg. Conf. Ser. in Appl. Math. SIAM, 1987, v + 78 pp, (P). [ISBN: 0-89871-213-0] Notes from the CBMS-NSF Conference on Probabilistic Methods in Combinatorics, Colorado, 1986. Topics include random graphs, the deletion method, the Gale-Berlekamp Switching Game, and the Lovász Local Lemma. LC

Probability, P. *Random Graphs '85.* Ed: Michał Karoński, Zbigniew Palka. Math. Stud., V. 144. North-Holland (US Distr: Elsevier Science), 1987, ix + 354 pp, $97.25 (P). [ISBN: 0-444-70265-2] Papers from the Second International Seminar on Random Graphs and Probabilistic Methods in Combinatorics, Poznań, August 1985. LC

Probability, T(15-16: 1, 2), S*, L*.** *Introduction to Probability.* J. Laurie Snell. Random House, 1988, xix + 517 pp, $27. [ISBN: 0-394-34485-5] Well-written text at the post-calculus level. Dominant feature is its extensive use of computer programs,

written in True BASIC, as a pedagogical tool to simulate and analyze chance experiments. Another distinguishing feature is the inclusion of numerous interesting historical remarks. Presents discrete and continuous probability in a separate but parallel manner so text could be used for a discrete probability course. Concludes with a chapter on finite Markov chains. RSK

Probability, S(15-16), L. *Applied Problems in Probability Theory.* E. Wentzel, L. Ovcharov. Transl: Irene Aleksanova. MIR (US Distr: Imported Pub), 1986, 432 pp, $13.95. [ISBN: 0-8285-3322-9] Revision of the 1983 Russian edition. Problems related to various fields, including electrical and radio engineering; data transmission; information systems; reliability, preventative maintenance, repair and accuracy of apparatus; transport; and health services. Each chapter begins by sketching (without proofs) some of the theory of probability; the bulk of each chapter is a collection of problems and their solutions which make use of the theory previously presented. An impressive collection of problems. LW

Elementary Statistics, T*(13-14: 1, 2). *Statistics, Fourth Edition.* James T. McClave, Frank H. Dietrich II. Dellen, 1988, xx + 1014 pp. [ISBN: 0-02-379260-4] Expanded and updated revision of the authors' 1985 *Third Edition* (*Second Edition,* TR, December 1982). Major change is the increased use of computer output throughout the text in both examples and exercises, allowing a change in emphasis from computations to interpretations. Other significant changes include an increased emphasis on statistical quality control, and a collection of computer exercises based on a data set available from the publisher. RSK

Elementary Statistics, T(13: 1). *Statistics Explained: Basic Concepts and Methods.* R. Kapadia, G. Andersson. Math. & Its Applic. Halsted Pr, 1987, 234 pp, $36.95. [ISBN: 0-470-20966-6] Presupposes no college mathematics. The usual topics, calculator tips, and Minitab used selectively. Many newspaper clippings showing statistics as it is used daily. Only birth data set and random numbers table are included. MS

Elementary Statistics, S(13). *How to Think about Statistics.* John L. Phillips, Jr. WH Freeman, 1988, xiv + 198 pp, $9.95 (P); $17.95. [ISBN: 0-7167-1923-1; 0-7167-1922-3] Revision of the 1982 *Second Edition* of the author's book *Statistical Thinking* (1973 *First Edition,* TR, August-September 1974). Presents the logic behind the major ideas in statistical thinking. Most chapters include several realistic situations, called sample applications, where the reader is asked to make a decision about what to do. Answer section includes both "a good choice" and "possible misinterpretations" for each. RSK

Statistics, T(14-16: 1, 2). *Probability and Statistical Inference, Third Edition.* Robert V. Hogg, Elliot A. Tanis. Macmillan, 1988, ix + 658 pp. [ISBN: 0-02-355810-5] For students with some calculus background. This new edition has more material on exploratory data analysis, more graphs, and more real applications. (*First Edition,* TR, May 1977.) FLW

Statistics, P, L.** *The Collected Works of John W. Tukey, Volume V, Graphics: 1965-1985.* Ed: William S. Cleveland. Stat./Prob. Ser. Wadsworth, 1988, lxiv + 464 pp, $44.95. [ISBN: 0-534-05102-2] Latest volume of a projected series covering Tukey's many and diverse contributions to statistics (see TRs, May 1985 of Series and *Volume I;* February 1986 of *Volume II;* and June-July 1987 of *Volumes III and IV*). Contains 18 papers, including some previously unpublished, documenting the inventiveness of his pioneering work in this area. Includes comments by the editor on each of the papers. RSK

Statistics, S(17), P, L. *Introduction to Optimization Methods and their Application in Statistics.* B.S. Everitt. Chapman & Hall, 1987, vii + 88 pp, $25 (P); $49.95. [ISBN: 0-412-29680-2; 0-412-27210-5] This short book can be broken into two sections. The first part discusses nonlinear mathematical programming algorithms independent of context, both direct and gradient methods. The second part contains several applications to statistical analysis. Good bibliography, a few problems, no solutions. SM

Statistics, T(16-18). *Linear Least Squares Computations.* R.W. Farebrother. Stat.: Textbooks & Mono., V. 91. Marcel Dekker, 1988, xiii + 293 pp, $79.75. [ISBN: 0-8247-7661-5] Prerequisite: sound knowledge of a computer programming language to implement the algorithms on computer. Matrix arithmetic introduced as required. Several aspects of least squares method examined; canonical expressions related to traditional expressions. Allows readers to select variants most suitable for their needs; eliminates error analysis; adopts an empirical approach. MS

Statistics, P. *Statistical Estimation in Large Parameter Spaces.* A.W. van der Vaart. CWI Tract V. 44. Math Centrum, 1988, 205 pp, Dfl. 31.60 (P). [ISBN: 90-6196-329-X] Theoretical monograph concerned with estimation in situations intermediate between parametric and nonparametric models. Obtains bounds on the asymptotic behavior of estimators, and shows how to construct efficient estimators for certain cases. RSK

Statistics, P, L. *Selected Tables in Mathematical Statistics, V. 11.* Ed: R.E. Odeh, J.M. Davenport. AMS, 1988, xi + 371 pp, $46. [ISBN: 0-8218-1911-9] Tables of one-sided and two-sided upper equicoordinate percentage points of the central multivariate student t distribution in which there is a common variance estimate in the denominators of the variates, and the numerators are equicorrelated. Applications of tables and examples of their use are described in detail. MS

Computer Literacy, C. *Using OS/2.* Kris Jamsa. Osborne McGraw-Hill, 1988, xv + 758 pp, $19.95 (P). [ISBN: 0-07-881306-9] A "how-to" book for using MicroSofts' new multi-tasking microcomputer operating system, OS/2. A straightforward, no frills coverage of concepts, commands, and procedures to

make use of OS/2. Most useful as a reference to accompany OS/2's user manual. PS

Computer Literacy, L. *Encyclopedia of Microcomputers, Volume 1: Access Methods to Assembly Language and Assemblers.* Ed: Allen Kent, James G. Williams. Marcel Dekker, 1988, viii + 434 pp, $160. [ISBN: 0-8247-2700-2] This is the first of a proposed ten-volume work of about 500 articles in about 5000 pages. This volume covers access methods to assembly language and assemblers. It has a 1988 copyright, but the entry for Apple, for example, does not have the Mac Plus, SE, or II mentioned. Two very good articles are "Architecture of microprocessors" (by Robert C. Stanley), and "Assembly language and assemblers" (by Roger Flynn). RSF

Elementary Computer Science, T*(13: 1). *A Second Course in Computer Science with Modula-2.* Daniel D. McCracken, William I. Salmon. Wiley, 1987, xiii + 474 pp, $34.10. [ISBN: 0-471-63111-6] Strongly emphasizes the use of abstract data types, recursion, and program readability. Makes effective use of Modula-2 to illustrate the software engineering principles discussed in the text. Content consistent with most recent ACM guidelines for a CS-2 course. AO

Elementary Computer Science, T*(13: 1). *Pascal on the Macintosh: A Graphical Approach.* David A. Niguidula, Andries van Dam. Addison-Wesley, 1987, xxxiv + 686 pp, (P). [ISBN: 0-201-16588-0] For the CS1 course. Interactive graphics orientation throughout: graphics before programming, not the other way around. Uses Macintosh and Lightspeed Pascal and all their debugging and formatting facilities. Emphasizes problem solving, program development, graphical user interfaces. Over 300 illustrations and 500 exercises. Later chapters have advanced topics. DFA

Programming, T(13-14). *Pascal Plus Data Structures, Algorithms, and Advanced Programming, Second Edition.* Nell Dale, Susan C. Lilly. DC Heath, 1988, xx + 750 pp, $26 (P). [ISBN: 0-669-15284-6] Good introduction to data structures and advanced programming. Emphasis on algorithms, modularization, data encapsulation and abstraction, information hiding, and software engineering make it a possible choice for a second semester programming text. Includes a new chapter on verifying, debugging, and testing, and updates in problem sets, built-in data structures discussion, and priority queues since the *First Edition* (TR, February 1986). Has always had readable text and nice illustrations. Sorting algorithms, efficiency considerations, and searching get higher visibility than some texts. LB-E

Programming, S, P. *OS/2 Programmer's Guide.* Ed Iacobucci. Osborne McGraw-Hill, 1988, xxi + 1100 pp, $24.95 (P). [ISBN: 0-07-881300-X] Written by the leader of the IBM OS/2 design team, this book for programmers (using standard edition release 1.0) is in three parts: the first part includes background material; the second part covers the overall structure and functions of OS/2; and the third part applies concepts from part 2. Examples are written in assembler; appendices include function calls, error codes, and sample programs. Topics presented include dynamic linking, the application programming interface (API), multi-tasking, protected mode, memory management, device drivers, I/O subsystems, session management, and OS/2 commands and utilities. RSF

Programming, S?? *An Easy Course in Using the HP-28S.* John W. Loux, Chris Coffin. Grapevine Pub, 1988, 315 pp, $22 (P). [ISBN: 0-931011-18-3] Hard to believe that anyone who purchased an HP-28S would have any use for this simplistic manual. Some helpful hints and shortcuts, but most are too little, too late. MR

Programming, T(14-15: 1), S. *The C Trilogy: A Complete Library for C Programmers.* Eric P. Bloom. TAB Books, 1987, viii + 588 pp, $22.95 (P). [ISBN: 0-8306-2890-8] This is a primer and reference manual for C. In addition to simply teaching the language, it also contains a listing of dozens of useful C functions which it calls the C Toolbox Library. GMS

Programming, P, L*. *Computer Games I.* Ed: David N.L. Levy. Springer-Verlag, 1988, xiii + 456 pp, $45. [ISBN: 0-387-96496-7] Part of a collection of 50 papers on how computers play games of strategy, drawn from widely-scattered sources, assembled by an International Chess Master and specialist in the field. *Volume I:* backgammon (world champion beaten!); chess; checkers; scrabble. *Volume II:* bridge; go; poker; nine other games. These solid papers concern aspects including mathematical analysis, heuristic strategies, practical considerations, history. Bibliography, index. RB

Programming, P. *C: Power User's Guide.* Herbert Schildt. Osborne McGraw-Hill, 1988, x + 382 pp, $22.95 (P). [ISBN: 0-07-881307-7] A text for those who want to become extremely competent programmers in C. Covers many advanced and specialized topics such as windows, graphics, screen programming, communications, and I/O programming. For the technical specialist who must use C in everyday work environment. GMS

Programming, T(13: 1). *Illustrating Pascal.* Donald Alcock. Cambridge U Pr, 1987, vii + 184 pp, $12.95 (P). [ISBN: 0-521-33695-3] Complete introduction to standard Pascal. Presentation is pictorial: the book is entirely hand-printed, with text and diagrams attractively interwoven throughout. Pleasant reading. Few programming exercises. Includes bubble sort, Quicksort, stacks and queues, symbol-state tables, reverse Polish notation, shortest-route technique, double-linked rings, binary trees, hashing. DFA

Languages, S(13-15), L. *Modula-2: A Complete Guide.* K.N. King. DC Heath, 1988, xxi + 656 pp, (P). [ISBN: 0-669-11091-4] Designed to be used either as a reference or as a textbook. Based on the third, corrected edition of Wirth's *Programming in Modula-2* (TR, June-July 1983). AO

Languages, T(18), P. *Concurrent Prolog: Col-*

lected Papers. Ed: Ehud Shapiro. Ser. in Logic Program. MIT Pr, 1987, $65 set [ISBN: 0-262-19255-1]. *Volume 1*, xxix + 525 pp; *Volume 2*, xviii + 653 pp. A republication of 39 papers by authors from various countries on concurrent forms of the logic programming language Prolog. One of these languages (GHC) is the "system language" for the Japanese ICOT Fifth Generation Computer Project, and much of the material was developed while interacting at ICOT. After surveying several concurrent logic programming languages, the work focuses primarily on two (Concurrent Prolog and its subset, Flat Concurrent Prolog). Parallel algorithms, system programming, high- and low-level implementation techniques. Suitable for reference or for the comparative study of concurrent logic programming languages; can possibly be used for learning (Flat) Concurrent Prolog. RB

Algorithms, P. *Distributed Algorithms and Protocols*. Michel Raynal. Transl: Jack Howlett. Wiley, 1988, xi + 163 pp, $31.95 (P). [ISBN: 0-471-91754-0] Protocols for controlling distributed systems. Deals with important problems such as nontrivial exclusion, deadlock, termination, and management of information transfers. Includes analyses of the algorithms and proofs of correctness. RWN

Algorithms, P. *Searching Algorithms*. Jiři Wiedermann. Teubner-Texte zur Math., B. 99. BG Teubner, 1987, 123 pp, 13 M (P). [ISBN: 3-322-00433-3] "[A] monograph as well as a handbook giving the overview of both the state-of-the-art and trends in the field of searching algorithms," by a Czech author. Advanced hashing techniques; various types of search trees; multidimensional searching with degrees of matching (e.g., interval match queries); synthesizing efficient solutions of searching problems via class membership problems and dynamization. RB

Computer Systems, S, P, L. *The UNIX System Guidebook, Second Edition*. Peter P. Silvester. Springer-Verlag, 1988, xiv + 334 pp, $28.95 (P). [ISBN: 0-387-96489-4] An introduction to UNIX for those with some computer experience who are concerned with applications rather than with computer systems themselves. Getting started, file system, shells, system kernel, utilities, editing, text processing, compilers, commands, bibliography. The *Second Edition* is expanded, extensively rewritten. Tested on Seventh Edition Unix, System V and 4.2 BSD versions of UNIX; emphasizes System V. (*First Edition*, TR, August-September 1984.) RB

Computer Systems, P. *Hypercube Multiprocessors 1987*. Ed: Michael T. Heath. SIAM, 1987, xvii + 761 pp. [ISBN: 0-89871-215-7] A collection of 80 papers given at the Second Conference on Hypercube Multiprocessors. Multiprocessor architecture has necessitated re-evaluating traditional programming techniques, algorithms, and operating systems to take full advantage of a multiprocessor system's potential. Papers presented illustrate this point, ranging from performance studies on hypercubes to many areas of their application including combinatorial problems, scientific applications, numerical methods, and special emphasis on matrix computations and partial differential equations. PS

Computer Systems, P. *Simulating Computer Systems, Techniques and Tools*. M.H. MacDougall. Ser. in Comput. Syst. MIT Pr, 1987, 292 pp, $32.50. [ISBN: 0-262-13229-X] Written for computer and communication system designers who want to analyze performance of designs. Divided into two parts. First, an introduction to discrete-event simulation in SMPL, a C-based simulation language. Covers model building, implementation in SMPL, and output data analysis. A multiprocessor system model and an ethernet model are carefully created as examples. The second part describes the implementation of SMPL. C source code for the SMPL language is included. PS

Computer Systems, P. *Software Diversity in Computerized Control Systems*. Ed: U. Voges. Dependable Computing & Fault-Tolerant Syst., V. 2. Springer-Verlag, 1988, 216 pp, $49.50. [ISBN: 0-387-82014-0] One of the methods of achieving fault-tolerant computer systems is through the use of fault-tolerant software. This text reviews the methods of developing such software and looks at applications of these techniques to such areas as transportation and nuclear reactors. GMS

Computer Systems, T(14: 1), S. *The MC68000 Assembly Language and Systems Programming*. William Ford, William Topp. DC Heath, 1987, xx + 590 pp, $24. [ISBN: 0-669-16085-7] Material for the one-semester assembly language programming course, and much more: on high-level language runtime environment, data structures, I/O programming (includes libraries), exception processing, peripheral device interrupts. Examples, complete programs, exercises. Appendices provide complete instruction set details for the MC68000 and related microprocessors. Extensively class tested. DFA

Computer Systems, S(15-16), P, L. *UNIX Internals: A Systems Operation Handbook*. Myril Clement Shaw, Susan Soltis Shaw. TAB Books, 1987, xiii + 206 pp, $17.95 (P). [ISBN: 0-8306-2951-3] An elementary introduction to the inner workings of the UNIX operating system. Doesn't burden the reader with too many technical details. Recommended for the user without an extensive background in computer science who wants (or needs) to know more about how UNIX does its job. AO

Computer Systems, S(15-16), P, L. *UNIX System Programming*. Keith Haviland, Ben Salama. Intern. Comput. Sci. Ser. Addison-Wesley, 1987, xv + 339 pp, (P). [ISBN: 0-201-12919-1] An excellent introduction to the system call interface of the UNIX operating system as defined by the AT&T System V Interface Definition. Also discusses the standard subroutine libraries. AO

Computer Graphics, P, L. *Geometric Modeling: Algorithms and New Trends*. Ed: Gerald E. Farin. SIAM, 1987, xi + 399 pp. [ISBN: 0-89871-206-8] A collection of survey articles and research papers organized into four sections: Curves and

Surfaces; Solid Modeling and Multivariate Surfaces; Shape; and Intersection Algorithms. About 60% of the twenty-seven papers are based on presentations made at the SIAM Conference on Geometric Modeling and Robotics held in Albany, New York, July 15-19, 1985. AO

Computer Graphics, P. *GKS Theory and Practice.* Ed: P.R. Bono, I. Herman. EurographicSeminars. Springer-Verlag, 1987, x + 316 pp, $49. [ISBN: 0-387-18257-8] The Graphics Kernel System (GKS) is the first international graphics standard. It is a description of a device-independent graphics library. This book emphasizes issues that arise in the implementation of GKS in a specific environment. It contains nineteen papers covering concepts and architecture, algorithms, specification, certification, and programming language interfaces. AM

Computer Graphics, T(16-17: 1), P, L*. *An Introduction to Splines for use in Computer Graphics and Geometric Modeling.* Richard H. Bartels, John C. Beatty, Brian A. Barsky. Morgan Kaufmann, 1987, xiv + 476 pp, $38.95. [ISBN: 0-934613-27-3] An intuitive introduction to parametric spline curves and parametric, tensor-product spline surfaces emphasizing those topics most applicable to computer graphics. Assumes only a background in elementary calculus and linear algebra. AO

Computer Graphics, P. *Intelligent CAD Systems I: Theoretical and Methodological Aspects.* Ed: P.J.W. ten Hagen, T. Tomiyama. EurographicSeminars. Springer-Verlag, 1987, xiv + 360 pp, $59.50. [ISBN: 0-387-18281-0] Eighteen papers contributed at an April 1987 workshop in The Netherlands. The title subject encompasses knowledge-based and expert CAD systems. Future workshops will concern implementational issues and practical experiences and evaluations. DFA

Computer Graphics, P. *Theoretical Foundations of Computer Graphics and CAD.* Ed: R.A. Earnshaw. NATO ASI Ser. F, V. 40. Springer-Verlag, 1988, xx + 1241 pp, $125. [ISBN: 0-387-19506-8] Proceedings of a 1987 institute on the analysis and exposition of the theoretical foundations of computer graphics and CAD. Topics include data structures, design, image processing, drawing algorithms, theory and formal methods, modelling and surfaces, image generation and reconstruction, hardware architectures. RM

Theory of Computation, S(18), P. *Lecture Notes in Computer Science-294: STACS 88.* Ed: R. Cori, M. Wirsing. Springer-Verlag, 1988, ix + 404 pp, $33.30 (P). [ISBN: 0-387-18834-7] This book contains reprints of the 34 papers presented at the Fifth Annual Symposium on the Theoretical Aspects of Computer Science, Bordeaux, France, February 1988. The papers address research topics in four major areas: complexity of algorithms, formal language theory, graph theory, and semantics. GMS

Artificial Intelligence, S(15-17), L*. *Perceptrons: An Introduction to Computational Geometry, Expanded Edition.* Marvin Minsky, Seymour Papert. MIT Pr, 1988, xv + 292 pp, $12.50 (P). [ISBN: 0-262-63111-3] Reprint of the 1972 printing (complete with its handwritten corrections overwritten on original 1969 typesetting), expanded with a contemporary prologue providing a twenty year retrospective: "We found that little of significance had changed since 1969, when the book was first published. ... The issues that were then obscure remain obscure today because no one yet knows how to tell which of the present discoveries are fundamental and which are superficial." LAS

Artificial Intelligence, T?(17-18: 1), P. *Computational Complexity and Natural Language.* G. Edward Barton, Jr., Robert C. Berwick, Eric Sven Ristad. MIT Pr, 1987, xii + 335 pp, $24.95. [ISBN: 0-262-02266-4] Study of computational complexity theory studied as a classification tool more powerful than weak generativity (classifying by the sets of strings capable of generation) for the analysis of grammatical formalisms. Analysis of agreement and ambiguity, lexical functional grammar, dictionary and morphological analysis sytems, and generalized phrase structure grammar (GPSG). RM

Artificial Intelligence, P. *Neural Computers.* Ed: Rolf Eckmiller, Christoph v.d. Malsburg. NATO ASI Ser. F, V. 41. Springer-Verlag, 1988, xiii + 566 pp, $112. [ISBN: 0-387-18724-3] A collection of 50 papers and 4 group reports covering topics such as neural network architecture, learning and memory, fault tolerance, pattern recognition, and motor control in brains versus neural computers. Intended as a reference book for research in the field of neural computers and in "the endeavor to transfer concepts of brain function and brain architecture to the design of self-organizing computers with neural net architecture." Includes a general reference list of introductory books on neural computers. PS

Artificial Intelligence, P. *Natural Language Generation Systems.* Ed: David D. McDonald, Leonard Bolc. Symbolic Computat. Springer-Verlag, 1988, xi + 388 pp, $37. [ISBN: 0-387-96691-9] Ten extended papers on computer systems for producing natural language output, concerning topics such as discourse theory, mechanical translation, deliberate (as opposed to spontaneous) writing, and revision. RB

Artificial Intelligence, P. *Real-Time Object Measurement and Classification.* Ed: Anil K. Jain. NATO ASI Series F, V. 42. Springer-Verlag, 1988, viii + 407 pp, $79.50. [ISBN: 0-387-18766-9] Proceedings of a workshop at Raretea, Italy, September 1987, focusing on robotics-related computer vision. 25 papers, including hardware implementation of specific vision algorithms, a complete vision system for object tracking and inspection, three-camera stereo for future measurement, object recognition by neural network, integration of CAD and vision systems, pyramid architectures in computer vision. RB

Computer Science, T(15-18: 1). *A Commonsense Approach to the Theory of Error Correcting Codes.* Benjamin Arazi. Ser. in Comput. Syst. MIT Pr, 1988, x + 208 pp, $22.50. [ISBN: 0-262-01098-

4] Presents this theory without much of the higher-level mathematics (finite fields, primitive polynomials) except in Appendices. Purpose is to provide an intuitive feel for the material, using basic linear algebra, exclusive-OR gates, and linear feedback shift registers. Includes most basic concepts in an easy-to-read format with concise descriptions. No problem-solving. LB-E

Computer Science, P. *Lecture Notes in Computer Science-296: Trends in Computer Algebra.* Ed: R. Janßen. Springer-Verlag, 1988, v + 197 pp, $21.80 (P). [ISBN: 0-387-18928-9] Proceedings of an international symposium sponsored by IBM Germany, Bad Neuenahr, May 1987, intended to provide a forum for researchers and users of computer algebra. Eleven papers and two abstracts: languages and systems; symbolic computations; computing in algebraic structures; applications. RB

Computer Science, P. *The Craft of Software Engineering.* Allen Macro, John Buxton. Intern. Comput. Sci. Ser. Addison-Wesley, 1987, xii + 380 pp, (P). [ISBN: 0-201-18488-5] A text for practitioners in the area of software design and development. It addresses such practical "how-to" issues as managing software development, estimating cost and timescale, team structure, estimating quality, and acceptance testing. This text is appropriate for experienced professionals who wish to improve their practical, technical, and management skills. GMS

Computer Science, S(17-18). *Lambda-Calculus, Combinators, and Functional Programming.* G.E. Revesz. Tracts in Theoret. Comput. Sci., V. 4. Cambridge U Pr, 1988, x + 181 pp, $29.95. [ISBN: 0-521-34589-8] The lambda calculus is the formal mathematical basis for the programming language LISP. There has been a renewed interest in this branch of mathematics because of its new applications to functional programming, denotational semantics, and artificial intelligence. This text introduces the reader to this area of mathematics and how it can be used in computer science. GMS

Computer Science, P, L. *Annual Review of Computer Science, Volume 2, 1987.* Ed: Joseph F. Traub, *et al.* Annual Reviews, 1987, ix + 565 pp, $39. [ISBN: 0-8243-3202-4] Sixteen survey papers reviewing recent developments in artificial intelligence, fault-tolerant computing, software engineering, and theoretical computer science. AO

Computer Science, P. *Mathematics in Signal Processing.* Ed: T.S. Durrani, *et al.* IMA Conf. Ser., V. 12. Clarendon Pr, 1987, xvi + 677 pp, $115. [ISBN: 0-19-853613-5] Proceedings of a conference at Bath, United Kingdom, September 1985. Application of advanced analytical techniques from mathematical sources such as approximation theory, number theory, linear algebra, optimization theory, time series analysis. 46 research papers on signal analysis and modelling; spectral analysis, inverse problems; image reconstruction; numerical algorithms and architectures; adaptive techniques. RB

Computer Science, P. *The Characteristics of Parallel Algorithms.* Ed: Leah H. Jamieson, Dennis Gannon, Robert J. Douglass. MIT Pr, 1987, x + 440 pp, $27.50. [ISBN: 0-262-10036-3] Contains reprints of papers presented at a workshop on Taxonomy of Parallel Algorithms held in Santa Fe, New Mexico, January 1987. There are 16 papers divided into three sections—classification of parallel algorithms, concurrency issues, and impact of parallelism on software development. GMS

Computer Science, P, L. *Research Directions in Object-Oriented Programming.* Ed: Bruce Shriver, Peter Wegner. Ser. in Comput. Syst. MIT Pr, 1987, vii + 585 pp, $40. [ISBN: 0-262-19264-0] Sixteen research papers describing current research in object-oriented programming in the areas of langauges, environments, databases, and theory. Preliminary versions of these papers were presented at a workshop held at the IBM T.J. Watson Research Center in June 1986. Early drafts of about half of these papers were published in the October 1986 issue of *SIGPLAN Notices.* AO

Computer Science, P. *Lecture Notes in Computer Science-292: The Munich Project CIP, Volume II: The Program Transformation System CIP-S.* CIP System Group. Springer-Verlag, 1987, viii + 522 pp, $40 (P). [ISBN: 0-387-18779-0] Formal descriptions of an implementation of a transformational programming system, a wide-ranging system in which the programmer can work at a variety of levels with automatic assistance in specification, program synthesis, code generation, verification, modification, and documentation. RWN

Computer Science, P. *Radon and Projection Transform-Based Computer Vision: Algorithms, A Pipeline Architecture, and Industrial Applications.* J.L.C. Sanz, E.B. Hinkle, A.K. Jain. Ser. in Inform. Sci., V. 16. Springer-Verlag, 1988, viii + 123 pp, $39. [ISBN: 0-387-18396-5] Describes the hardware and algorithms used in a real-time, projection-based machine vision system. Also describes several applications to industrial visual inspection problems. AO

Computer Science, S(18), P. *Abstract Interpretation of Declarative Languages.* Ed: Samson Abramsky, Chris Hankin. Books in Comput. Sci. Halsted Pr, 1987, 284 pp, $64.95. [ISBN: 0-470-20971-2] Devoted to a new research area in computer science called abstract interpretation, a technique which involves the application of non-standard semantics to the compilation of programs. The algorithms developed are applied to declarative languages such as PROLOG. The book requires a great deal of sophisticated knowledge in such advanced areas as compiler design and denotational semantics. GMS

Computer Science, T(13). *An Introduction to Computer Science with Modula-2.* J. Mack Adams, Philippe J. Gabrini, Barry L. Kurtz. DC Heath, 1988, xix + 571 pp, (P). [ISBN: 0-669-12171-1] A well-organized and well-written introduction to computer science. Presents an excellent problem solving methodology and clearly illustrates it in each example throughout the text. Also places emphasis on

the role of documentation in programming—a topic needing emphasis during introductory courses. Covers the usual other material: recursion (using several examples with debugging traces), iteration, elementary data structures, control structures. Modular programming's importance is both stressed and justified—the choice of Modula-2 reinforces this. Other topics include simple sorts, some algorithm analysis, abstract data types, numerical methods, and concurrency. PS

Computer Science, P. *Lecture Notes in Computer Science-291: Graph-Grammars and Their Application to Computer Science.* Ed: H. Ehrig, *et al.* Springer-Verlag, 1987, viii + 609 pp, $52.10 (P). [ISBN: 0-387-18771-5] Proceedings of a 1986 workshop on graph grammars (grammar methods for specifying and generating graphs and rewriting mechanisms). Contains both tutorial papers and research contributions, with applications to parallelism, concurrency, modelling, complexity, etc. RM

Applications, T(15-17), P, L. *Fuzzy Mathematical Techniques with Applications.* Abraham Kandel. Addison-Wesley, 1986, xiv + 274 pp. [ISBN: 0-201-11752-5] A systematic introduction to the mathematical theory and applications of imprecise relationships. Fuzzy sets transcend classical probability in scope: imprecision (fuzziness) includes not only sets but also probabilities, statistics, relations, and logics. Applications conclude with a fuzzy relational data base as a model of a human expert. Includes 1000-item bibliography. LAS

Applications. *Mathematical Modelling: Classroom Notes in Applied Mathematics.* Ed: Murray S. Klamkin. SIAM, 1987, xiv + 338 pp, (P). [ISBN: 0-89871-204-1] A collection of the articles which have appeared in the "Classroom Notes in Applied Mathematics" section of *SIAM Reviews* since 1975 together with an extensive list of supplemental references organized by subject area. AO

Applications (Economics), P. *Econometrics of Planning and Efficiency.* Ed: Jati K. Sengupta, Gopal K. Kadekodi. Adv. Stud. in Theoret. & Appl. Econom., V. 11. Kluwer Academic, 1988, ix + 193 pp, $66.50. [ISBN: 90-247-3602-1] A collection of ten papers concerning stochastic planning, a combination of econometrics with aspects of probability and operations research. Three parts: methodology, planning models, and analysis of efficiency. SM

Applications (Electrical Engineering), T(16-17: 1). *An Introduction to Signal Detection and Estimation.* H. Vincent Poor. Texts in Elect. Engin. Springer-Verlag, 1988, x + 549 pp, $58. [ISBN: 0-387-96667-6] Provides good introduction to (or review of) hypothesis testing while pre-supposing a working knowledge of applied probability and random processes. Excellent exercises at the end of each chapter require both theoretical and application understanding. Separates subject into problems of detection and estimation. Fine explanations. LB-E

Applications (Engineering), T(15-16: 2, 3). *Advanced Engineering Mathematics.* Stanley I. Grossman, William R. Derrick. Harper & Row, 1988, xviii + 1176 pp. [ISBN: 0-06-042534-2] In six mostly independent parts for a year-long course covering ordinary and partial differential equations, vector calculus, linear algebra, numerical methods, and complex variables. In addition to the usual applications in electrical circuits and vibratory motion, topics in economics, engineering, and physics are included. Not wordy, but very clearly written. Stress on the need for solving problems. Instructor's manual available. JK

Applications (Engineering), T(15: 1, 2). *Advanced Engineering Mathematics.* Michael D. Greenberg. Prentice-Hall, 1988, xii + 946 pp. [ISBN: 0-13-010505-8] Traditional material on applied mathematics for students in engineering or science. Twenty chapters in five (mostly independent) parts. For one or more semesters. Prerequisites are calculus and at least part of a course in ordinary differential equations. Written to be readable by students. Useful as a reference and for self study within and without a course. Except for the finite-difference solution of partial differential equations, topics in numerical analysis are omitted. Drill-type exercises. Some problems on applications, others extend the text. Answers to selected exercises. Solutions manual available. JK

Applications (Fluid Mechanics), P, L. *Annual Reviews of Fluid Mechanics, Volume 20, 1988.* Ed: John L. Lumley, Milton van Dyke, Helen L. Reed. Annual Reviews, 1988, 551 pp, $34. [ISBN: 0-8243-0720-8] Seventeen survey papers on various topics (e.g., fractals, surf-zones, sand transport, digital image processing) together with cumulative author and title indices to the entire twenty volume series. LAS

Applications (Information Theory), P. *Self-Organization and Associative Memory, Second Edition.* Teuvo Kohonen. Ser. in Inform. Sci., V. 8. Springer-Verlag, 1988, xv + 312 pp, $39.50 (P). [ISBN: 0-387-18314-0] An exposition of mathematical aspects of patterning processing. This edition includes new chapters on neural computing and optical associative memories. (*First Edition,* TR, November 1984.) SG

Applications (Medicine), P. *Mathematics and Computer Science in Medical Imaging.* Ed: Max A. Viergever, Andrew Todd-Pokropek. NATO ASI Ser. F, V. 39. Springer-Verlag, 1988, viii + 546 pp, $98. [ISBN: 0-387-18672-7] A summary of the proceedings of a NATO Advanced Study Institute held in Il Ciocco, Italy, September 21-October 4, 1986. Covers fundamental ideas in medical imaging and the associated mathematical problems. Aimed at physicists, engineers, computer scientists, mathematicians, and informed clinicians working in the area of medical imaging. Includes eleven introductory and tutorial papers and nineteen more specialized papers. RH

Applications (Physical Science), P. *Multiprocessing in Meteorological Models.* Ed: G.-R. Hoffmann, D.F. Snelling. Topics in Atmospheric & Oceanic Sci. Springer-Verlag, 1988, xvi + 438 pp,

$70. [ISBN: 0-387-18457-0] Predicting the weather is harder than von Neumann imagined, but parallel processing should extend the range and accuracy of weather models. Contains 28 papers from two workshops, including an excellent article on "Computer Requirements for Atmospheric Modelling." BC

Applications (Physics), S(13), L. *Seven Ideas that Shook the Universe.* Nathan Spielberg, Bryon D. Anderson. Wiley, 1987, viii + 263 pp, $22.95. [ISBN: 0-471-85974-5] A descriptive, non-mathematical approach to seven themes in physics: Copernican astronomy, Newtonian mechanics, the energy concept, entropy, relativity, quantum theory, and conservation principles and symmetries. Describes the historical and philosophical context in which the ideas arose, and discusses the relationship between these ideas and our understanding of the physical world. The quantitative material is presented mostly in the form of graphs and diagrams, and very few formulas are included. RH

Applications (Physics), S(14-16), L. *The Expanding Universe.* Arthur Eddington. Cambridge U Pr, 1987, xxiii + 128 pp, $12.95 (P). [ISBN: 0-521-34976-1] Re-issue in the Cambridge Science Classics series of a set of public lectures delivered by Eddington in 1932 in Cambridge, Massachusetts, in which he attempts to convey in personal, non-mathematical terms a cosmological model linking relativity with quantum theory in a manner consistent with Hubble's observations. Lively, enlightening, closely reasoned—a vivid mix of science and art. LAS

Applications (Physics), S*(16-17), L.** *Elementary Particles and the Laws of Physics: The 1986 Dirac Memorial Lectures.* Richard P. Feynman, Steven Weinberg. Cambridge U Pr, 1987, x + 110 pp, $12.95. [ISBN: 0-521-340004] The first two lectures in a series established at Cambridge University to commemorate P.A.M. Dirac. Feynman speaks on the reason for antiparticles, Weinberg on the final laws of physics. Super exposition by master interpreters of modern physics. LAS

Applications (Physics), P. *Probabilistic Methods in Mathematical Physics.* Ed: Kiyosi Itô, Nobuyuki Ikeda. Academic Pr, 1987, vii + 441 pp, $59.50. [ISBN: 0-12-375660-X] Proceedings of the Taniguchi International Symposium in Katata and Kyoto, 1985. MU

Applications (Physics), T(13: 1), L. *Beyond the Mechanical Universe: From Electricity to Modern Physics.* Richard P. Olenick, Tom M. Apostol, David L. Goodstein. Cambridge U Pr, 1986, xiv + 574 pp, $27.95. [ISBN: 0-521-30430-X] The second volume of a college physics text which accompanies the acclaimed PBS series *The Mechanical Universe ... and Beyond,* a Caltech-based project to "infuse new life in freshman physics." This volume corresponds to the last half of the television series: electricity; magnetism; relativity; particles; quantum mechanics. An offering institution would presumably provide a "live, flesh-and-blood college physics teacher" and a laboratory component. Text may be used alone without the television series; teacher supplements available. RB

Applications (Physics), P. *Functional Integrals and Collective Excitations.* V.N. Popov. Mono. on Math. Physics. Cambridge U Pr, 1987, viii + 216 pp, $49.50. [ISBN: 0-521-30777-5] Begins with an introduction to functional integral techniques in equilibrium statistical physics and the expression of partition functions and Green functions in terms of functional integrals. Later sections cover superfluid Bose systems, plasma and superfluid Fermi systems, crystals, heavy atoms, and model Hamiltonians. MU

Applications (Physics), S, P, L*. *Nonlinear Functional Analysis and its Applications, IV: Applications to Mathematical Physics.* Eberhard Zeidler. Transl: Juergen Quandt. Springer-Verlag, 1988, xxiii + 975 pp, $129. [ISBN: 0-387-96499-1] The fourth of a five-volume set that includes historic background of modern ideas and attempts to bridge the gaps between mathematicians and physicists. Definite tilt of author to the notion that mathematicians should know a good deal about the assumptions underlying mathematical models results in a book in which physics and mathematics do appear unified. Numerous quotations that introduce each chapter are themselves fascinating reading; an impressive, erudite, readable book. AWR

Applications (Simulation), T(17-18), S, P. *Neural and Brain Modeling.* Ronald J. MacGregor. Academic Pr, 1987, xii + 643 pp, $89. [ISBN: 0-12-464260-8] Part I: an extensive review (presumably the first) of the literature in neural and brain modeling, surveying engineering, mathematical, and computer-based modeling methodology. Part II: a computer model (54 FORTRAN programs) which simulates the dynamics of neurons and neuronal networks, accompanied by over 200 pages of instructions, examples and exercises, for access by students, scholars, and researchers. RB

Reviewers

RJA: Richard J. Allen, St. Olaf; DFA: David F. Appleyard, Carleton; RB: Richard Brown, Carleton; LB-E: Lisa Brown-Ewing, Macalester; JNC: Judith N. Cederberg, St. Olaf; LC: Laura Chihara, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; RSF: Robert S. Fisk, St. Olaf; SG: Steven Galovich, Carleton; GG: George Gilbert, St. Olaf; BH: Bruce Hanson, St. Olaf; RH: Rhonda Hatcher, St. Olaf; RK: Roger Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; AM: Alan Magnuson, St. Olaf; SM: Steve McKelvey, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; MR: Matthew Richey, St. Olaf; AWR: A. Wayne Roberts, Macalester; GMS: G. Michael Schneider, Macalester; PS: Paul Schoening, St. Olaf; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MS: Myriam Steinback, Macalester; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; LW: Lynne Walling, St. Olaf; FLW: Frank L. Wolf, Carleton; MZ: Michael Zielinski, St. Olaf.

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, DC 20036